Boris Vassilev

Studies on **Proteins** Influencing **Cancer Progression** and Regulating **Endocytic Lipid Trafficking**

Academic Dissertation

Research conducted at
University of Helsinki, Faculty of Medicine, Department of Anatomy
and Minerva Foundation Institute for Medical Research

Thesis supervisor: Elina Ikonen
Thesis pre-examiners: Petri Auvinen and Simon Anders

Doctoral Programme in Biomedicine
Thesis committee: Sampsa Hautaniemi and Sergey Kuznetsov

ISBN 978-951-51-3098-3 (paperback)
ISBN 978-951-51-3099-0 (PDF)
Unigrafia Oy, Helsinki 2017

Contents

List of Figures

Glossary

3-HMG-CoA	3-hydroxy-3-methylglutaryl CoA
DIC	differential interference contrast
D-MEM	Dulbecco's modified Eagle's medium
DiI-LDL	1,1′-dioctadecyl-3,3,3′,3′-tetramethyl-indocarbocyanine-perchlorate – labeled LDL
EEA1	early endosome antigen 1
E-MEM	Eagle's minimum essential medium
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ER	endoplasmic reticulum
ErbB2	receptor tyrosine-protein kinase erbB-2
ESCRT	endosomal sorting complexes required for transport
FAK	focal adhesion kinase
FBS	fetal bovine serum
HMGR	3-hydroxy-3-methylglutaryl-coenzyme A reductase
IDOL	inducible degrader of the LDL-receptor
LAPTM4B	lysosomal-associated transmembrane protein 4B
LDL	low-density lipoprotein
LDLR	low-density lipoprotein receptor
LE	late endosome
LXR	sterol-responsive nuclear liver X receptor
MVB	multi-vesicular body
NDRG1	N-myc downstream-regulated gene 1
NPC	Niemann-Pick disease, Type C
NPC1	NPC intracellular cholesterol transporter 1
NPC2	NPC intracellular cholesterol transporter 2
nSREBP	nuclear SREBP
PBS	phosphate-buffered saline
PM	plasma membrane
PRA1	prenylated Rab acceptor protein 1
S1P	Site-1 protease
S2P	Site-2 protease
SCAP	SREBP cleavage-activating protein
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLO	streptolysin O
Sphk1	sphingosine kinase 1
Src	proto-oncogene tyrosine-protein kinase Src
SRE	sterol regulatory element DNA sequence
SREBP	sterol regulatory element-binding protein
StARD3	StAR-related lipid transfer protein 3
TIRF	total internal reflection fluorescence
VLDL	very low-density lipoprotein

Original Publications

I. Vilja Pietiäinen, **Vassilev B**, Blom T, Wang W, Nelson J, Bittman R, Bäck N, Zelcer N, Ikonen E: NDRG1 functions in LDL receptor trafficking by regulating endosomal recycling and degradation. *Journal of Cell Science* 2013 Sep 1; 126(Pt 17): 3961–71.

II. **Boris Vassilev**, Sihto H, Li S, Hölltä-Vuori M, Ilola J, Lundin J, Isola J, Kellokumpu-Lehtinen PL, Joensuu H, Ikonen E: Elevated levels of StAR-related lipid transfer protein 3 alter cholesterol balance and adhesiveness of breast cancer cells: potential mechanisms contributing to progression of HER2-positive breast cancers. *American Journal of Pathology* 2015 Apr; 185(4): 987–1000.

III. **Boris Vassilev**, Louhimo R, Ikonen E, Hautaniemi S: Language-agnostic reproducible data analysis using literate programming. *PLoS ONE* 2016 11(10): e0164023.

Abstract

Endocytosis is a form of active cellular transport by which cells internalize molecules from the extracellular environment. This thesis studies proteins involved in endocytosis that influence the progression of cancers in humans. Two of the three articles summarized here employ cell biological experiments to elucidate the role of the proteins N-myc downstream-regulated gene 1 (NDRG1) and StAR-related lipid transfer protein 3 (StARD3) in the regulation of endocytic trafficking. The third article describes a novel data analysis method and its application to a large data set of breast cancer patients.

We studied the effects of NDRG1 protein depletion on cellular cholesterol content and distribution, low-density lipoprotein receptor (LDLR) localization and turn-over, and the morphology of endosomal organelles. We found that depletion of the NDRG1 protein leads to reduced abundance of LDLR on the plasma membrane (PM) and as a result, reduced low-density lipoprotein (LDL) uptake. When NDRG1 is depleted, LDLR accumulates in multi-vesicular bodies (MVBs), and while LDLR ubiquitination is increased, its degradation is decreased. The PM levels of LDLR and LDL uptake are rescued upon co-depletion of the inducible degrader of the LDL-receptor (IDOL). Our findings identify NDRG1 as regulator of MVB formation and LDLR trafficking.

We also found that elevated levels of StARD3 protein alter the cholesterol balance and adhesiveness of breast cancer cells. We studied the effects of StARD3 overexpression in cells, as well as the association of StARD3 protein levels with cancer progression markers in Finnish breast cancer patients. We found that in MCF-7 cells, stable StARD3 overexpression altered the morphological features and the cholesterol balance of the cells. In the breast cancer tumor samples from patients, high StARD3 protein levels associated with *ErbB2* (receptor tyrosine-protein kinase erbB-2) amplification. High protein levels also associated with proto-oncogene tyrosine-protein kinase Src (Src) activation and increased 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) transcript levels. Our findings suggest that elevated StARD3 levels contribute to breast cancer aggressiveness.

Finally, a systematic approach to data analysis was developed and demonstrated by applying it to an existing data set of breast cancer patients. High-throughput molecular level data and clinical data were integrated in the study. This revealed a coamplification of *NDRG1* and the lysosomal-associated transmembrane protein 4B gene (*LAPTM4B*), encoding an endosomal protein that regulates ceramide trafficking, and discovered new genes potentially coregulated with *LAPTM4B*.

1 Introduction

Cells take up and recycle nutrients and signaling molecules by endocytosis. Endocytosis is an active vesicular transport mechanism. Vesicles are formed at the cell surface as membrane invaginations that are pinched off the plasma membrane and internalized. These early endosomes contain cargo like low-density lipoprotein (LDL). Their membrane includes proteins like cell surface receptors, for example the LDL receptor. Cell receptors can be degraded in the lysosome or recycled back to the cell surface, providing a regulating mechanism.

Cancer cells escape the tight control of the organism over growth, differentiation, and cell cycle. Oncogenes cause cancer, either by gene mutation or high gene expression. *HER2*, is an oncogene encoding the cell surface receptor protein Her2, overexpressed in a subtype of cancer. Her2-positive breast patients can be treated with an antibody against the cell surface receptor, but the mechanism of action is complex and not fully understood, and relapse is commonly observed.

HER2 is at the core of an amplified genomic region that contains several other genes. A commonly co-amplified gene encodes StARD3, a late endosomal membrane protein, with a cholesterol-binding transmembrane domain and a cholesterol-binding pocket domain protruding into the cytosol. Our preliminary experiments showed that the cellular levels of the StARD3 protein can influence cell growth. Further studies showed that this effect is partially due to a deregulation of the endocytic pathway.

This thesis studies how the deregulation of endocytosis can influence cancer progression. The focus of the experimental work was on two genes: StAR-related lipid transfer protein 3 (StARD3), and N-myc downstream-regulated gene 1 (NDRG1). Our results suggest that StARD3 contributes to breast cancer aggressiveness, probably by decoupling *de novo* cholesterol synthesis from external growth stimuli. While NDRG1 was shown to regulate LDL receptor trafficking in particular, its role in multi-vesicular body formation suggests a more general function as a regulator of cell surface receptor degradation.

To provide a systemic view on the regulation of endocytosis in the context of cancer progression, we analyzed data containing genome wide DNA amplification and mRNA levels, as well as clinical patient data from The Cancer Genome Atlas (TCGA). We identified a module of genes that are regulated by the same micro-RNA as *LAPTM4B*, a gene encoding a protein involved in lipid transport in the late endosome. In addition, we demonstrated how to organize and present the data analysis in order to improve its repeatability and reproducibility.

This book provides an introduction to the topics of cellular cholesterol balance, endocytic transport, receptor signalling, the role of cholesterol in cancer, and the analysis of biological data. A review of the most relevant scientific literature provides a context for the questions and findings reported in the original publications. The experimenal methods summarize the tools used to conduct this research.

2 Review of the Literature

2.1 Cellular cholesterol balance

In humans, as in all animals, cholesterol in lipid bilayers of the cell regulates their fluidity. Cholesterol is also a precursor for the synthesis of steroid hormones. Steroids are molecules based on a tetracyclic ring system (fig. 1). In humans, most steroids act as hormones: for example testosterone, androsterone (androgens), estrone and estradiol (estrogens), and adrenocortical hormones. The two sources of cholesterol are *de novo* synthesis and dietary uptake.

Triacylglycerols, cholesterol, and other lipids obtained from the diet are taken up by the intestine and carried by the blood stream as chylomicrons. Triacylglycerides are hydrolyzed, and the cholesterol-rich chylomicron remnants are taken up by the liver; triacylglycerols and cholesterol are released back to the blood stream in the form of very low-density lipoprotein (VLDL). Again, triacylglycerides are hydrolyzed, and the remnants, intermediate-density lipoproteins, are either taken up by the liver or converted to low-density lipoprotein (LDL). The main role of LDL is to transport cholesterol to peripheral tissues and regulate cholesterol synthesis there. Feedback mechanisms in the cell sense low cholesterol levels and activate cholesterol synthesis (Ikonen 2008).

2.1.1 Cholesterol synthesis

The finding by Konrad Bloch that acetate is a precursor of cholesterol provided the first clue to elucidating the pathway for the biological synthesis of cholesterol in animals (Bloch 1965). In fact, all 27 carbon atoms of cholesterol are derived from acetyl CoA, and mevalonate and squalene are key intermediates (section 2.1). First, acetyl

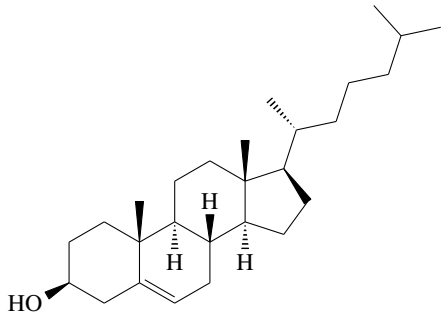


Figure 1: Chemical structure of the cholesterol molecule. The molecule is not soluble in water. The tetracyclic steroid nucleus of cholesterol is planar and rigid. When embedded within the lipid bilayer, the hydroxyl group interacts with the polar heads of the phospholipids and sphingolipids in the membrane, while the steroid and the hydrocarbon chain interact with the fatty acid chains of the other lipids, influencing membrane fluidity, thickness, and membrane protein affinity.

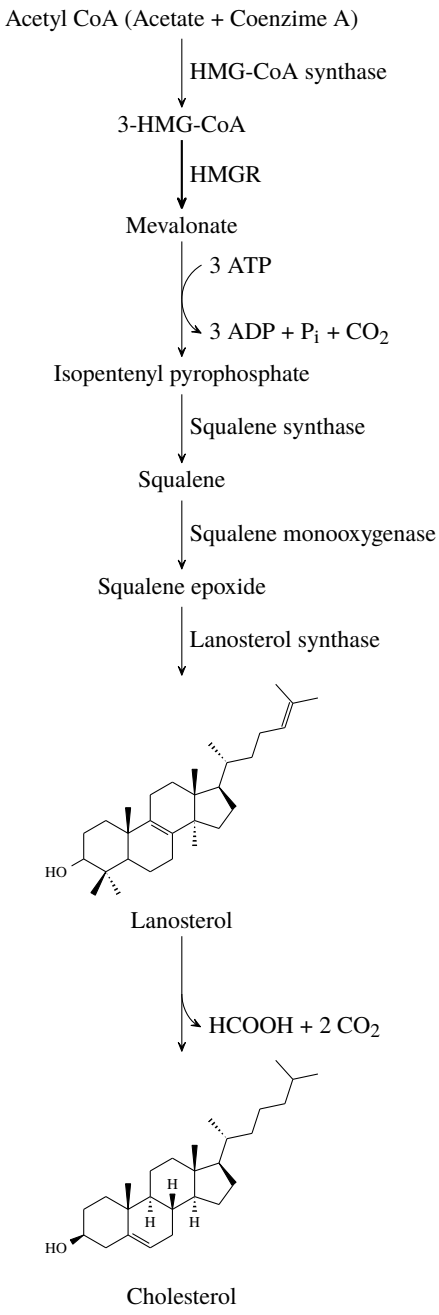


Figure 2: Major steps in cholesterol biosynthesis. Each arrow represents a chemical reaction that consists of multiple intermediate steps. The committed step in cholesterol biosynthesis, the production of mevalonate, is shown with a thick arrow. In the last step, lanosterol is converted to cholesterol in 19 intermediate reactions that require nine different enzymes. Three methyl groups of lanosterol are oxidized and removed as formic acid and two carbon dioxide molecules.

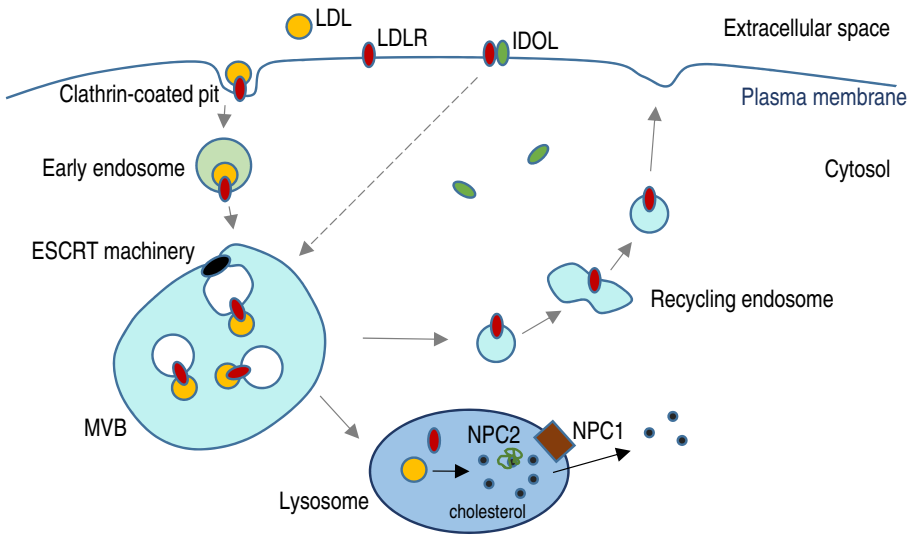


Figure 3: Overview of LDL uptake, and LDL receptor recycling and degradation. LDL receptor accumulates in clathrin-coated pits, where it binds LDL. Early endosomes with their cargo mature to multivesicular bodies. This process is regulated by the ESCRT machinery – **e**ndosomal **s**orting **c**omplexes **r**equired for **t**ransport. As the acidity increases, LDL dissociates from the LDL receptor. LDL receptor is recycled back to the plasma membrane via recycling endosomes. IDOL, **i**nducible **d**egrader **o**f the **L**DL receptor, is recruited by LDL receptor to the plasma membrane. IDOL provides an ESCRT-mediated pathway for LDL receptor to the multivesicular bodies and to lysosomes that is independent of clathrin-coated pits or caveolae. In the lysosome, LDL receptor is degraded, while LDL particles are broken down, cholesteryl esters are hydrolyzed, and free cholesterol exits the lysosome in a process regulated by the NPC1 and NPC2 intracellular cholesterol transporter proteins.

CoA, acetoacetyl CoA, and water form 3-hydroxy-3-methylglutaryl CoA (3-HMG-CoA). 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) is the catalyst of the reaction:

$$\text{3-HMG CoA} + 2\,\text{NADPH} + 2\,\text{H}^+ \xrightarrow{\text{HMGR}} \text{mevalonate} + 2\,\text{NADP}^+ + \text{CoA}$$

This reaction is irreversible. It is the committed step in cholesterol formation, and HMGR is the rate-determining enzyme.

Mevalonate is converted to isopentenyl pyrophosphate in three consecutive reactions that require ATP. Squalene is synthesized from six molecules of isopentenyl pyrophosphate. It is oxidized to squalene epoxide in a reaction that requires molecular oxygen and is catalyzed by squalene monooxygenase. Squalene epoxide is cyclized to lanosterol by lanosterol synthase. Lanosterol is then converted to cholesterol.

2.1.2 Cholesterol uptake from LDL

Free cholesterol, with its polar OH group, would be confined to the surface of lipoproteins like LDL. Instead, the cholesterol in LDL is in the form of cholesteryl esters: the OH

group is esterified with a fatty acid, to make the molecule non-polar. LDL is internalized into cells by clathrin-mediated endocytosis. On the plasma membrane (PM), low-density lipoprotein receptor (LDLR) associates with clathrin-coated pits and promotes the intake of LDL and its delivery to early endosomes (fig. 3). From early endosomes, LDL particles are transported to multi-vesicular bodies (MVBs) and late endosomes (LEs). The low endosomal pH dissociates the LDLR from the LDL particles and LDL is trafficked via LEs to lysosomes. There, LDL cholesterol esters are hydrolyzed to free cholesterol by acid lipase (Sugii et al. 2003). Exit of cholesterol from the LEs is regulated by NPC intracellular cholesterol transporter 1 (NPC1) and NPC intracellular cholesterol transporter 2 (NPC2) (Carstea et al. 1997; Loftus et al. 1997). NPC1 is needed to recruit Rab8a to LEs, and Rab8a enhances the motility and segregation of recycling endosomes from LEs and lysosomes (Kanerva et al. 2013).

On the PM, inducible degrader of the LDL-receptor (IDOL) blocks LDL association with the LDLR in clathrin-coated pits, and regulates LDLR uptake and degradation (Scotti et al. 2011). IDOL employs the MVB pathway to sort LDLR to the lysosome for degradation. Endosomal sorting complexes required for transport (ESCRTs) are evolutionary conserved proteins found even in archaea that control other cellular functions: receptor signaling, cytokinesis, autophagy, polarity, migration, miRNA activity and mRNA transport (Zaremba-Niedzwiedzka et al. 2017). More recent findings suggest that the endosomal sorting complexes required for transport (ESCRT) pathway is required for IDOL-mediated LDLR degradation, while LDLR internalized through clathrin-mediated endocytosis is preferentially recycled to the PM (Scotti et al. 2013).

2.1.3 Regulatory mechanisms

The rate of cholesterol synthesis is highly responsive to the cellular level of cholesterol. The feedback mechanisms are primarily mediated by changes in the amount and activity of HMGR (fig. 4). Cholesterol generated from LDL hydrolysis within the lysosome is responsible for suppressing 3-HMG-CoA activity through the sterol regulatory element-binding protein (SREBP) pathway. By inhibiting the SREBP pathway, LDL suppresses the transcription of *LDLR*, along with other proteins involved in cholesterol uptake and synthesis (Goldstein and M. S. Brown 2009).

Newly synthesized SREBP is inserted into the ER membrane and is bound to the SREBP cleavage-activating protein (SCAP). (fig. 4, ①). SCAP is a sensor for sterols: when cells are depleted of cholesterol, SCAP escorts SREBP from the ER to the Golgi, where the two proteases S1P and S2P sequentially cleave SREBP to release nSREBP (fig. 4, ②). The cleaved nSREBP translocates to the nucleus where it activates SREs. (fig. 4, ③). SRE is regulated by SREBPs SREBP-1a, SREBP-1c, and SREBP-2. In the presence of sterols, SRE inhibits HMGR mRNA synthesis rate. Of the three SREBPs, SREBP-1a activates genes mediating the synthesis of cholesterol, fatty acids, and triglycerides, while SREBP-1c activates genes mediating fatty acid synthesis and glucose metabolism. SREBP-2 preferentially activates cholesterol synthesis enzymes HMG-CoA synthase, HMGR, and squalene synthase, as well as LDL receptor (Horton et al. 2002).

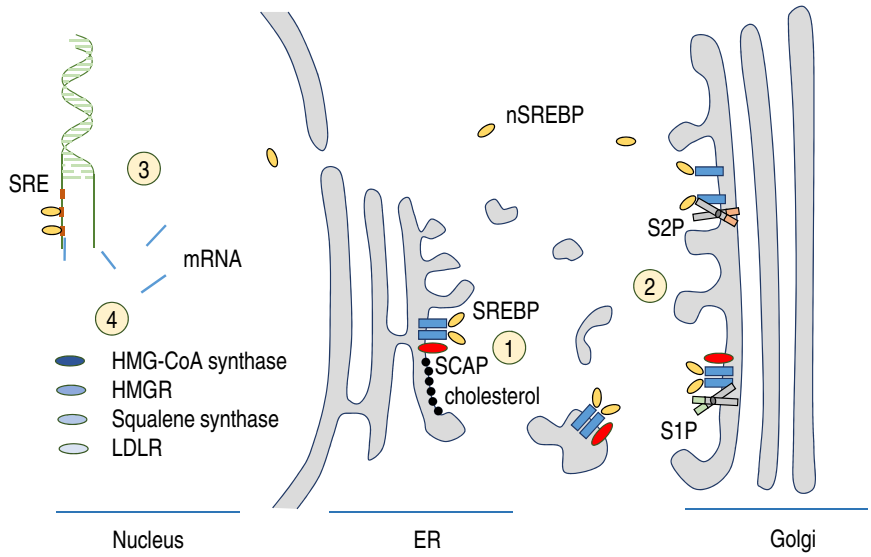


Figure 4: Cellular cholesterol balance is regulated through the SREBP pathway. The numbers designate the site for each consecutive step: ① – transport of SREBP from the endoplasmic reticulum (ER) to the Golgi; ② – cleavage of SREBP to nuclear SREBP (nSREBP) by Site-1 protease (S1P) and Site-2 protease (S2P); ③ – activation of sterol regulatory element DNA sequences (SREs); and ④ – transcription of the relevant mRNAs. See the text for details.

Feedback mechanisms complementary to the SREBP pathway have also been described. For example, sterol-responsive nuclear liver X receptor (LXR) maintains cholesterol homeostasis both through promotion of cholesterol efflux and suppression of LDL uptake. LXR inhibits the LDLR pathway through transcriptional induction of IDOL by triggering the ubiquitination of LDLR (Zelcer et al. 2009).

The rate of translation of the HMGR mRNA is inhibited by nonsterol mevalonate products like isoprenoid (Nakanishi et al. 1988). Furthermore, the membrane domain of HMGR senses high levels of the derivatives of cholesterol and mevalonate in the membrane, leading to the rapid degradation of HMGR. Oxysterols such as 25-hydroxycholesterol and 24,25-dihydrolanosterol stimulate ubiquitination of lysine residues in the membrane domain of HMGR. This triggers extraction of HMGR from the ER membrane and delivery to proteasomes for degradation (Jo and Debose-Boyd 2010).

2.1.4 Methods for studying cellular cholesterol

Biologically relevant molecules can be studied by using fluorescent or radioactive labels, or by using molecules that bind them, influencing their biological activity or allowing the studied molecule to be visualized. For example, dipyrromethene difluoride-cholesterol is a fluorescent cholesterol analogue that has been successfully used to visualize the cellular

distribution and accumulation of cholesterol (Hölttä-Vuori et al. 2008, 2016). However, experimental methods that employ a cholesterol analogue are limited by the fact that the fluorescent moieties tend to perturb the physical properties of the lipid.

The incorporation of radioactively labeled cholesterol precursors (for example acetate) into cholesterol can be used to measure the rate of *de novo* cholesterol synthesis in cells without changing the physical and chemical properties of cholesterol (Heino et al. 2000). The isolation of subcellular fractions can be used to study the localization of radioactively labeled cholesterol. However, the available experimental methods for purifying subcellular compartments limit the sensitivity of such studies. Pure fractions are exceedingly difficult to obtain.

Another approach is to use biologically active molecules that target cholesterol in the cell or the PM. Streptolysin-O (SLO) is a toxin that binds to cholesterol-containing membranes to assemble to curved rod structures that penetrate the apolar domain of the membrane bilayer, generating trans-membrane pores (Bhakdi et al. 1985). Treating cells with SLO can be used as an indirect measure of the free cholesterol content of the PM: for example, the release of cytosolic lactate dehydrogenase (LDH) upon SLO treatment can be used as a readout for pore formation (Pimplikar et al. 1994).

Filipin is a fluorescent antibiotic that binds to cholesterol, but not to esterified sterols: it can be used for detecting free cholesterol. Filipin binding permeabilizes the cells, so it cannot be used for live cell imaging, and the molecular basis for filipin binding is not completely understood (Maxfield and Wüstner 2012). Still, filipin staining intensity correlates well with the amounts of unesterified cholesterol in cells and can be used as a tool for quantitative imaging (Linder et al. 2007).

2.2 Endocytic pathway

Endocytosis is the transport of macromolecules, particles, and even other cells from the PM into the cell. Phagocytosis involves the intake of large particles as microorganisms or dead cells via phagosomes. It is mainly done by specialized white blood cells, phagocytes: macrophages, neutrophils, and dendritic cells.

All eucaryotic cells continually ingest some of their PM in the form of small endosomal vesicles. Conversely, the surface and the volume of the cell are maintained by continually removing membrane and cargo by exocytosis. Receptor-mediated endocytosis provides a mechanism for selective concentration and uptake of extracellular ligands. Cholesterol uptake through LDL by LDLR-mediated endocytosis is the major route for cholesterol into the cell. The formation of clathrin-coated pits is one mechanism for the formation of endosomal vesicles. These are specialized regions of the PM, with a diameter in human epithelial cells typically within a 120 nm to 150 nm range (McMahon and Boucrot 2011). After cargo selection, clathrin-coated pits invaginate into the cell and pinch off to form clathrin-coated vesicles. Clathrin-coated vesicles rapidly shed their clathrin coat and fuse with early endosomes.

Another pathway for endocytosis is through caveolae, invaginations in the PM of most cell types. Caveolae are flask-shaped, with a diameter between 70 nm and 120 nm. Based on their size and lipid content, caveolae are a subset of lipid rafts: specialized membrane

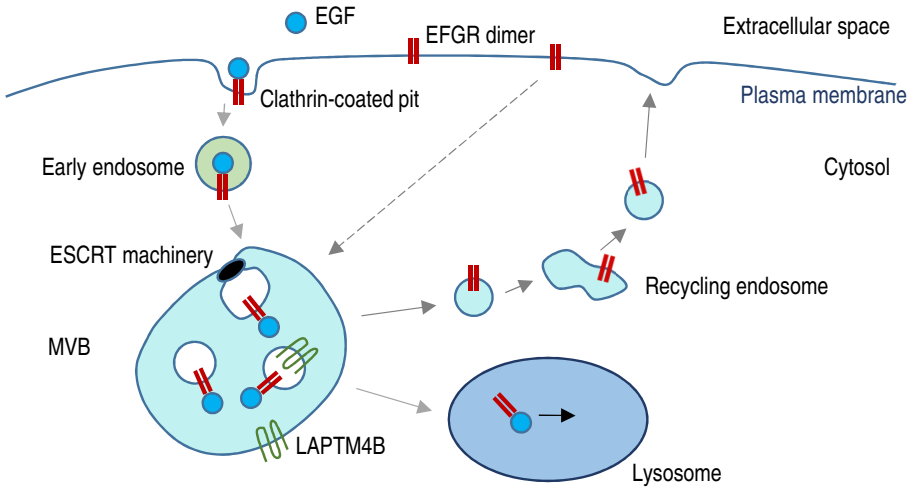


Figure 5: EGF receptor uptake, recycling, and degradation. Unlike LDL receptor, EGF receptor binds its ligand EGF before localizing to clathrin-coated pits. It is transported through multivesicural bodies, with the help of the ESCRT machinery, to the lysosome, where both the receptor and its ligand are degraded, or it is recycled back to the plasma membrane. There is another, ligand-independent pathway for EGF receptor uptake regulated by lysosomal-associated transmembrane protein 4B (LAPTM4B) that does not depend on clathrin-coated pits, and which targets the receptor for degradation. See the text for details.

microdomains enriched in sphingolipids and cholesterol. Both clathrin-coated pits and caveolae function in endocytosis, signal transduction, and receptor recycling (Reeves et al. 2012). Dynamin-2 is a protein located along the neck of the invaginations and promotes caveolae budding (Yao et al. 2005). Cross-linked GPI-anchored proteins move into caveolae and become endocytosed. Caveolae are enriched in signaling lipids such as ceramide, and ceramide is produced in caveolae in response to interleukin-1 beta, an important mediator of the inflammatory response (Simons and Vaz 2004).

2.2.1 Degradative functions

Early endosomes mature to MVBs that are moved along microtubules and become LEs, either by fusing with each other or with existing LEs. LEs communicate with the trans-Golgi network via transport vesicles, for example by obtaining the proteins that convert the LEs into lysosomes (Keller and Simons 1997).

After internalization into early endosomes, cell surface receptors may have different fates. Both LDLR and epidermal growth factor receptor (EGFR) can be either recycled to the PM through recycling vesicles and tubules, or targeted to lysosomes for degradation (fig. 3 and fig. 5). EGFRs bound to their ligand, epidermal growth factor (EGF), accumulate in clathrin-coated pits and after uptake are degraded in the lysosomes along with EGF (Sigismund et al. 2008).

2.2.2 Endosomal ceramide

The multivesicular endosome membrane can be sorted to the lysosome for degradation, or secreted in the form of exosomes. Ceramide is required for the segregation of cargo, and for the transfer of exosome-associated domains into the lumen of the endosome (Trajkovic et al. 2008). In LEs, LAPTM4B interacts with ceramide to remove it from LEs (Blom et al. 2015). In LAPTM4B-depleted A431 cells, ceramide accumulates in LEs, and stable overexpression of LAPTM4B in A431 cells lowers cellular ceramide content. Silencing LAPTM4B in KPL-4 breast cancer cells protects them from chemotherapy-induced caspase-3 activation. Recruitment of sphingosine kinase 1 (Sphk1) to sphingosine-enriched endocytic vesicles and the generation of sphingosine-1-phosphate facilitates membrane trafficking along the endosomal pathway. Furthermore, Sphk1 inhibitors induce Sphk1-dependent fusion of endosomal membranes (Young et al. 2016). This leads to accumulation of enlarged LEs, a phenotype that could be rescued by activation of ceramide synthase and the restoration of autophagy.

Interestingly, LAPTM4B is also involved in the trafficking and degradation of EGFR (Tan et al. 2015a,b). LAPTM4B inhibits EGF-stimulated intraluminal sorting of EGFR and thus its degradation by promoting the ubiquitination of an ESCRT subunit. Thus, LAPTM4B and ESCRTs regulate EGFR signaling independently of clathrin-mediated endocytosis (fig. 5).

2.2.3 StARD3 and late endosomes

The *StARD3* gene is overexpressed in human breast carcinomas and exhibits a homology with the steroidogenic acute regulatory protein STAR (Moog-Lutz et al. 1997). It has a cholesterol-binding trans-membrane domain that targets it to LEs, and a cytosolic cholesterol binding pocket (Tsujishita and Hurley 2000; Alpy et al. 2005; Alpy and Tomasetto 2006; van der Kant et al. 2013). The StARD3 protein is transported to the LE via the PM, and its cholesterol binding pocket is necessary for late endocytic dynamics and exit of free cholesterol from lysosomes (Zhang et al. 2002). The cholesterol binding activity of StARD3 is involved in the association of LEs with actin filaments and the actin-dependent movement of LEs (Hölttä-Vuori et al. 2005). In NPC1-deficient chinese hamster ovary cells in which cholesterol transport to the mitochondrial inner membrane is not affected by the lack of NPC1, StARD3 knockdown results in decreased cholesterol transport to the inner membrane and reduces mitochondrial cholesterol levels. Thus, StARD3 might provide a pathway for the egress of cholesterol from endosomes to mitochondria (Charman et al. 2010). This observation is supported by the finding that StARD3 and its close paralogue StARD3NL are involved in the formation of contacts between LEs and the endoplasmic reticulum: they interact with ER-anchored VAP-A and VAP-B proteins, allowing ER-endosome contact formation (Wilhelm et al. 2016). The presence and importance of contact sites between ER and mitochondria are well established (Rowland and Voeltz 2012); these recent findings highlight the role of StAR-related lipid transfer protein 3 (StARD3) in contacts between endosomes, and the ER and mitochondria.

2.3 EGF receptor signaling

Receptor tyrosine kinases are a class of enzyme-linked cell surface receptors. These transmembrane proteins have their ligand-binding domain on the outside of the cell. In the membrane, two or more receptor chains come together to form a dimer or an oligomer that is activated by ligand binding. Receptor tyrosine kinases then phosphorylate each other at multiple tyrosines in the cytoplasmic tails, which transmits the activation signal.

2.3.1 ErbB receptor family

EGFR belongs to the ErbB family of receptor tyrosine kinases implicated in cancers; its ligand is EGF. The main responsibility of the EGF signal is to stimulate proliferation. Of the four family members, ErbB2 is the most potent oncoprotein. It is the only protein in the ErbB family for which a high-affinity ligand has not been identified (Yarden and Sliwkowski 2001). ErbB2 forms dimers with the other family member receptors, creating docking sites for adapter proteins or enzymes that stimulate downstream pathways (fig. 6). Through induction of downstream signaling pathways, ErbB receptors enhance cancer cell survival and proliferation, migration and adhesion, tumorigenesis, as well as apoptosis inhibition (Karamouzis et al. 2007; Yarden and Sliwkowski 2001). For example, the PI3K/Akt/mTOR pathway regulates the cell cycle; the cSrc/FAK pathway regulates cell migration, adhesion, and angiogenesis; and the cSrc/STAT pathway regulates tumorigenesis.

Interestingly, cross-talk between EGF receptor signaling and cholesterol balance has been reported. For example, a family member, ERBB4, can undergo ligand-induced proteolytic cleavage to release a soluble intracellular domain that enters the nucleus to modify transcription and activate SREBP-2 (Haskins et al. 2015).

2.4 Cholesterol and cancer

The role of dietary cholesterol and plasma cholesterol levels in cancer progression has been extensively studied. For instance, in a Finnish nation-wide study, pre-diagnostic statin use was found to be associated with lowered risk of breast cancer death in a dose- and time-dependent manner (Murtola et al. 2014). In a Danish nationwide study, statin use in patients with cancer was associated with reduced cancer-related mortality in 13 types of cancer including breast cancer (Nielsen et al. 2012). In a mouse model of mammary tumor formation, dietary cholesterol accelerated and enhanced tumor formation. Moreover, plasma cholesterol levels were reduced during tumor development but not prior to its initiation (Llaverias et al. 2011).

The cellular homeostasis of cholesterol is also involved in cancer progression. For example, overexpression of *PPP1R1B-STARD3* fusion transcript in MKN-28 cells significantly increased cell proliferation and colony formation by increasing activation of PI3-kinase and AKT signaling (Yun et al. 2014).

Mutations in the tumor suppressor protein p53 have been shown to influence cancer progression by directly regulating *de novo* cholesterol synthesis. Mutant p53 can upregulate mevalonate pathway genes via SREBP transcription factors. In cell lines derived from

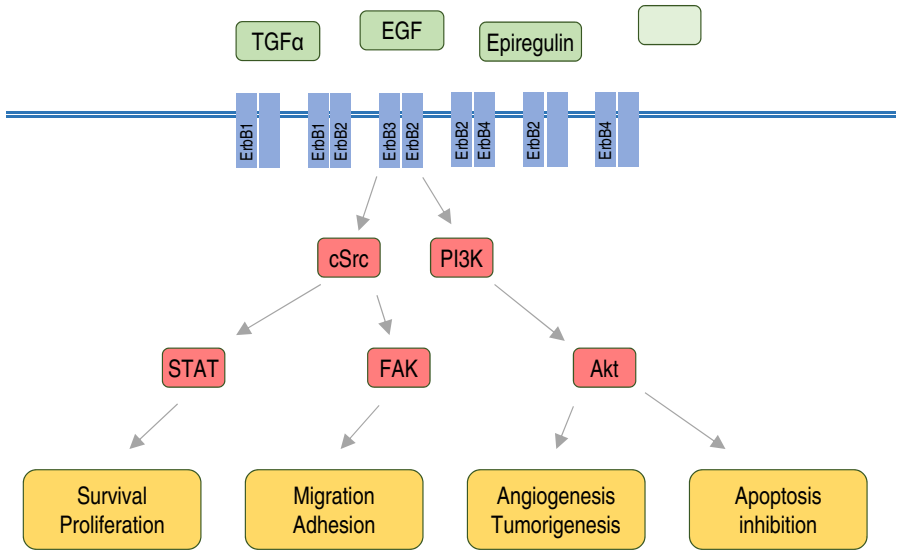


Figure 6: A simplified diagram of selected ErbB receptor family signaling pathways. Some of the most important ligands for the ErbB receptor family dimers are shown in green; there many other known ligands (the light green empty box). The known receptor dimers are shown in blue; if an EGF receptor is known to form dimers with multiple other receptor family members, the second box is left empty. The most important downstream signal transducers are shown in red, and the cellular processes that they regulate are shown in yellow.

metastatic breast tumors and grown in 3D culture, HMGR inhibitors mimicked the phenotypic effects of mutant p53 depletion. The p53 mutation correlated with the elevation of mevalonate pathway genes in human tumors (Freed-Pastor et al. 2012).

The *SQLE* gene encodes squalene monooxygenase. In a study of a panel of breast cancer cell lines, *SQLE* inhibition resulted in a decrease in cell viability, and in a noticeable increase in replication time. The effects depended both on gene copy number and on inhibitor dosage. These effects were observed only in the cell lines with detectable *SQLE* transcript, suggesting that *SQLE* can act as an oncogene in breast cancer (D. N. Brown et al. 2016).

Oxidosqualene cyclase is a lanosterol synthase, a post-squalenic enzyme of cholesterol biosynthesis. In a study employing a selective inhibitor of oxydosqualene cyclase, Ro 48-8071, enzyme inhibition had a strong anti-angiogenic effect by impairing endothelial cell adhesion and migration, and by blocking vessel formation in angiogenesis assays by specifically inhibiting AKT phosphorylation (Maione et al. 2015).

2.4.1 NDRG1 in cancer

The N-myc downstream-regulated gene 1 (NDRG1) protein is widely expressed in human tissues, mostly in epithelial cells, and is strongly upregulated under hypoxia (Lachat

et al. 2002). In vitro overexpression of NDRG1 in metastatic colon cancer cells reduced invasion; in mice, overexpression suppressed colon cancer metastasis (Guan et al. 2000). The expression of the NDRG1 protein in prostate cancer patient tumors was inversely correlated with overall survival, while in mice, it inhibited lung colonization of metastatic prostate cancer cells (Bandyopadhyay et al. 2003).

The levels of the NDRG1 protein were significantly reduced in patient breast tumors and metastases, while in vitro overexpression suppressed the invasiveness of breast cancer cells (Bandyopadhyay et al. 2004). In pancreatic cancer cells, NDRG1 overexpression lead to decrease in tumor growth and a reduction in tumor-induced angiogenesis (Maruyama et al. 2006). In colorectal cancer tissue, patients with lower *NDRG1* mRNA had a lower survival rate (Strzelczyk et al. 2009). In prostate cancer cell lines, NDRG1 recruited onto recycling endosomes, interacted with membrane bound Rab4aGTPase, and colocalized with transferrin (Kachhap et al. 2007).

The NDRG1 protein has also been specifically implicated in cellular cholesterol homeostasis. In a genome-wide expression analysis of Niemann-Pick disease, Type C (NPC) fibroblasts, *NDRG1* mRNA was elevated (Reddy et al. 2006). 1,1′-dioctadecyl-3,3,3′,3′-tetramethyl-indocarbocyanine-perchlorate – labeled LDL (DiI-LDL) uptake in HeLa was stimulated by knockdown of *NDRG1* (Bartz et al. 2009).

2.4.2 StARD3 in cancer

STARD3 is one of several genes coamplified with *ERBB2* and found in the same genomic region (17q12-q21). StARD3 knockdown results in decreased cell proliferation of overexpressing cancer cell lines (Kauraniemi and Kallioniemi 2006; Kao and Pollack 2006; Staaf et al. 2010). Its overexpression in ErbB2 overexpressing cell lines was required for the growth and survival of breast cancer cell lines (Sahlberg et al. 2013). Until very recently, however, very little was known about the function of the StARD3 protein, or the mechanism by which its co-amplification in ErbB2-positive breast cancers contributed to cancer progression.

2.5 Generation of knowledge from existing data

Applying high-throughput screening to experiments has generated a wealth of data such as nucleotide and protein sequences, protein crystal structures, gene-expression measurements, protein and genetic interactions and phenotype studies (Howe et al. 2008). This accumulation of publicly available data has created an opportunity and a need: generating biologically interesting results from existing “raw” data.

2.5.1 Studies of the TCGA database

The Cancer Genome Atlas (TCGA) is a collaboration between the National Cancer Institute and the National Human Genome Research Institute. TCGA has generated comprehensive, multi-dimensional maps of the key genomic changes in 33 types of cancer, including breast cancer. The data is publicly available.

TCGA has provided a comprehensive molecular portrait of human breast tumors by analyzing genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays (Koboldt et al. 2012). These data has been used by many bioinformatics studies. The study by Grieb et al. (2014) is just one example of utilizing the TCGA database. Evaluation of mRNA expression and copy number variation data revealed that the MDM2 Binding Protein is commonly overexpressed in breast cancer and increased transcript of gene amplification significantly correlated with reduced breast cancer patient survival.

2.5.2 Sharing and reproducing computational research

Automatic computers have enabled us to conduct research in a radically novel way (Dijkstra 1989). In computational research, performing an experiment amounts to instructing a machine to carry out arithmetic and logical operations automatically. A high level free text description of the instructions carried out by the machine is usually enough to explain the rationale and the logic of the experiment to other researchers. To instruct the computer to carry out the experiment, a computational scientists must provide the "program": a definition of the experiment in a formal language. Computer architectures vary, and the instructions that define the same experiment on two different machines will be different. Until the last ten years of last century, computers were rare and very different from each other. It was usual that a computer program had to be rewritten for every machine it ran on.

However, the situation has been changing. There are fewer computer architectures still developed and sold, and existing architectures have converged to some degree. General-purpose computers have become affordable and ubiquitous. At the same time, the idea that computer programs can and should be freely available is slowly gaining acceptance. It is increasingly easy to do computational research using only free*software, which has further encouraged the development and standardization of free tools for computational research. As a result, it is easier than ever to write a computer program, share it, and have others run it with minimal effort.

Currently, sharing and reproducing computational research can be achieved by following simple guiding principles and adhering to straightforward rules and best practices (fig. 7). For example, the computational analysis should be understandable to someone unfamiliar with the project, and everything that was done once should be easy to repeat, either exactly as is, or after improving or correcting it (Noble 2009). This can be done by keeping track of how every result was produced, avoiding manual data manipulations, providing full access to the original data and the analyzes, as well as explaining the design and purpose of the experiments (Sandve et al. 2013; Wilson et al. 2014).

Many of the rules, best practices, and guiding principles in computational research are well known in the field of software engineering: version control, self-documenting code, automated documentation generation, automated tests, scripting, open source. However, in software engineering, the final product is usually the software itself. In computational research, the software is just a means to obtaining a scientifically relevant result. The

*"Free" as in "free speech", not as in "free beer".

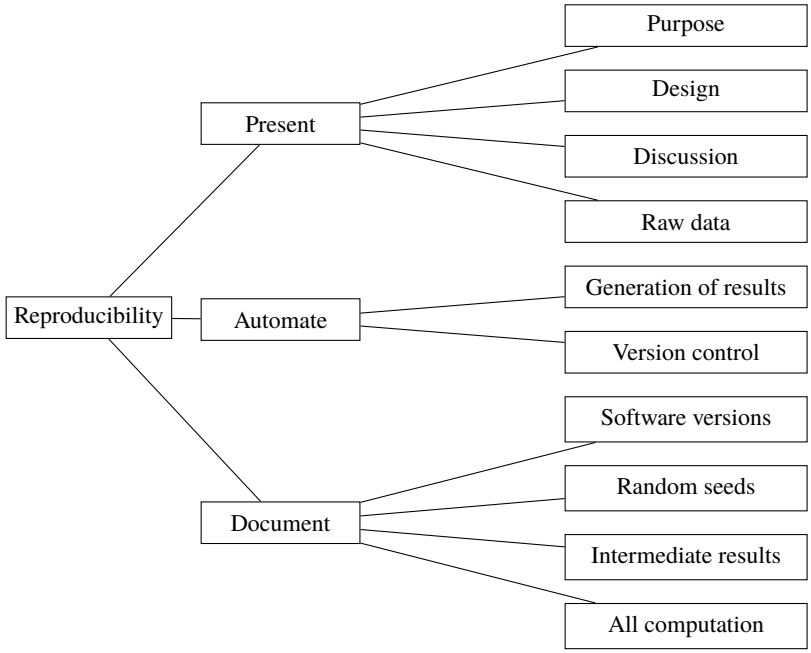


Figure 7: Summary of the best practices and rules for reproducible computing. For example, to improve *Reproducibility*, *document* all *intermediate results*, and *present*, to the human reader, the *purpose* and the *design* of the data analysis.

electronic notebook is the alternative to an experimental notebook – it contains the software and the results obtained, and can be shared with others to inspect, reproduce, or modify (Shen 2014). Existing electronic notebook implementations include Sweave and Knitr for R (Leisch 2002), IPython for Python (Pérez and Granger 2007), and Jupyter for almost any widely used programming language. A different approach is to embed all software within the PDF files published by scientific journals: it is invisible in a conventional PDF reader, but a special tool can extract the source code from the PDF file (Kitchin 2015).

For larger, more complex computational analysis projects, a lab notebook might not be well-suited. Frameworks for large-scale data acquisition and integration aim to facilitate projects involving many people and data sources by providing better tools for collaboration and documentation, along with libraries for acquiring data. These frameworks incur some unavoidable overhead to install and use (Ovaska et al. 2010; Börnigen et al. 2015).

The work by Donald Knuth on literate programming (Knuth 1983, 1984) has influenced the concept of the lab notebook. The literate programming tool Noweb (Ramsey 1994, 2008) is a generalization and simplification of the original work by Knuth. Sweave, the first widely used lab notebook software tool, is build upon Noweb.

3 Aims

The overall aim of this study was to gain insights into the functions of the endocytic lipid trafficking regulators NDRG1, StARD3, and LAPTM4B, and how they may contribute to cancer progression. The following three main questions were addressed in the course of the study:

- What is the role of NDRG1 in cellular lipid homeostasis and LDLR endocytic trafficking, and how is it relevant in cells expressing the protein abundantly?
- Does the cholesterol-binding protein StARD3 contribute to the progression of Her2-positive breast cancer and if so, what mechanisms are involved?
- What further clues to the role of these proteins in breast cancer progression can be extracted from an existing genome-wide breast cancer patient dataset?

4 Methods

Methods from different fields were employed: cell biology, biochemistry, statistics, and computational data analysis. For each method, the original publications in which it was used, as numbered in "Original Publications" on page vi, is given in parenthesis. The methods that were used in previously unpublished results presented in this thesis are marked with 'U'.

Importantly, only the materials and methods for the experiments performed by the author of this thesis summary are listed here; for full details on all materials and methods, see the original publications.

4.1 Cell biology

4.1.1 Cell culture (I, II, U)

All cell lines were cultured at 37 °C in a CO_2-conditioned, humidified incubator. A431 cells were cultured in Dulbecco's modified Eagle's medium (D-MEM) supplemented with 10 % fetal bovine serum (FBS), 0.5 mM L-glutamine and 100 IU/mL penicillin, 100 µg/mL streptomycin.

Oli-neu cells were cultured in D-MEM with B27 supplement on 50 µg/mL poly-D-lysine-coated plastic.

MCF-7 cells were cultured in Eagle's minimum essential medium (E-MEM) supplemented with 10 % FBS, 0.5 mmol/L L-glutamine, 0.01 mg/mL insulin, non-essential amino acids, and 100 IU/mL penicillin and 100 µg/mL streptomycin.

KPL-4 cells were cultured in D-MEM supplemented with 5 % FBS.

BT-474 cells were cultured in D-MEM supplemented with 10 % FBS, 1 mM sodium pyruvate, 4 mM L-glutamine, 0.01 mg/mL insulin, as well as 100 IU/mL penicillin and 100 µg/mL streptomycin.

4.1.2 Plasmid and siRNA transfections (I, II)

Plasmid DNA was introduced to cells either using Lipofectamine-LTX or Effectene as a transfection reagent, following the manufacturer's protocol. To inhibit gene expression, siRNAs were introduced to cells using the HiPerfect transfection reagent, following the manufacturer's protocol.

4.1.3 Lentiviral particle transduction (U)

Cells were plated on 48-well plates 24 h before transduction. At the time of transduction cells were 50 % confluent. Equal amounts of lentiviral particles (functional viral titer was determined using the p24 viral capsid protein) were added to a 1:1 mixture of polybrene and full medium, added to the cells, and incubated for 5 h before replacing with fresh full medium. After 24 h of incubation the selection antibiotic, 1 µg/µL puromycin, was added to the culture medium. The cells were cultured at 37 °C in a CO_2-conditioned, humidified incubator.

4.1.4 DiI-LDL binding and uptake (I)

Cells were incubated with DiI-LDL for 30 min at 4 °C for plasma membrane (PM) receptor binding and 20–30 min at 37 °C for internalization in serum-free medium prior to fixation with 4 % PFA and nuclear staining with DAPI.

4.1.5 Immunofluorescent staining (I, II, U)

For imaging of focal adhesions, cells were fixed and permeabilized with methanol for 5 min at −20 °C. For filipin staining, the cells were fixed, stained, and permeabilized with 0.05 % filipin in 10 % FBS in phosphate-buffered saline for 30 min at 37 °C.

Otherwise, cells were fixed with 4 % paraformaldehyde for 20 min at room temperature, and autofluorescence was quenched in 50 mmol/L ammonium chloride for 15 min at room temperature. Cells were permeabilized with 0.1 % Triton X-100 in phosphate-buffered saline (PBS) for 5 min and blocked with 10 % FBS for 30 min at room temperature. Primary and secondary antibody incubations were done in separate steps.

4.1.6 Cell imaging (I, II, U)

Red/green fluorescent images were acquired with an Olympus AX70 microscope using 10×, 20×, and 40× air objectives.

Filipin ultraviolet light images were acquired with the same microscope and with an inverted Olympus microscope with a Polychrome IV monochromator using a 10× air objective.

Confocal images were acquired with a Leica TCS SP2 AOBS confocal microscope with a 40× and 63× oil immersion objective.

Total internal reflection fluorescence (TIRF) images were acquired with a Nikon Eclipse Ti-E inverted microscope equipped with an argon laser and a 100× oil immersion objective with 1.49 numerical aperture. The corresponding wide-field images of filipin and StARD3-GFP were taken at a focal plane 1 µm above the TIRF plane.

Phase-contrast and green fluorescent images of live cells and fixed cells were acquired with a Zeiss AxioVert 200 microscope equipped with a Zeiss AxioCam HRc camera using a 10× air objective.

Differential interference contrast (DIC) images of live cells on plastic were taken with an inverted Olympus microscope with a 10× air objective.

4.2 Biochemistry

4.2.1 Immunoblotting (I, II, U)

Cell lysates were collected in a stringent denaturing lysis buffer (1 % SDS, 5 mM EDTA) containing phosphatase inhibitors (1 mM activated ortho-vanadate, 25 mM sodium fluoride) and protease inhibitors chymostatin, leupeptin, antipain and pepstatin A (CLAP), boiled for 5 min and passed through a needle to decrease viscosity. Total protein concentration in each sample was determined by a BioRad colorimetric protein assay. All samples were

adjusted to the same protein concentration before loading to a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel. After separation, the proteins were transferred to a nitrocellulose membrane and immunoblotted. Secondary antibodies were detected either with an Odyssey infrared scanner, or with enhanced chemiluminescence reagents and films by Amersham.

4.2.2 Lipid extraction and thin layer chromatography (I, II)

Lipids were extracted from cells using the Bligh and Dyer method (Bligh and Dyer 1959). Free cholesterol and cholesterol esters were resolved on high-performance thin layer chromatography plates using hexane, diethyl ether, and acetic acid in 80:20:1 ratio as the mobile phase. Total amounts were quantified by charring and comparing to a standard lipid mix separated on the same plate. Cholesterol synthesis and cholesterol efflux in cells fed with radioactively labeled lipid precursors was measured by separating the total lipids, then scraping the plate according to the migration of non-radioactive standards and detecting the scintillation counts.

4.2.3 Cell membrane permeabilization (II)

Cells were permeabilized in a cell membrane cholesterol-dependent manner using the Streptolysin O (SLO) toxin from *Streptococcus pyogenes*. After 30 min of cell membrane permeabilization at 37 °C the medium was collected and the cells were lyzed. The decrease in absorbance at 340 nm, resulting from the oxidation of NADH by LDH on addition of pyruvate to each sample was measured over 1 min. The slope of the absorbance curve, δ_M for the medium and δ_L for the lysate, provides a measure for the amount of LDH in a sample. For each medium-lysate pair, the permeability ρ of the membrane after SLO treatment was calculated as:

$$\rho = \frac{\delta_M}{\delta_M + \delta_L}$$

and is indicative of the amount of cholesterol in the cell membrane.

4.2.4 Cell surface biotinylation (I)

The complete assay was performed at 4 °C in a cold room. Cells were washed in PBS containing 0.02 mM $CaCl_2$ and 0.15 mM $MgCl_2$ (PBS+), cell surface proteins were biotinylated for 20 min on ice with 0.5 mg/mL biotin in PBS+, and unbound biotin was quenched by two incubations of 5 min in 0.1 M glycine and 0.3 % BSA in PBS+. Cells were then rinsed twice in PBS+ and lyzed in PBS containing 0.2 % SDS, 2 % NP-40, and protease inhibitors (CLAP). Lysates were collected and cleared by centrifugation, and an aliquot taken for the total protein amount. Biotinylated proteins from the rest of the lysate were precipitated with 50 µL streptavidin-agarose beads by rotation overnight. The beads were then washed before eluting the protein from the beads by boiling in 2× sample buffer; at this point the samples are analyzed by immunoblotting. To quantify the degradation of a cell-surface protein, cells were chased for 0 h, 2 h and 5 h in complete medium at 37 °C before cell surface biotinylation.

4.2.5 Quantitative Real-Time RT-PCR (II)

Samples were homogenized in RLT buffer from Qiagen, and total RNA was isolated with the RNeasy Mini Kit from Qiagen, according to the manufacturer's protocols. 1 µg of the isolated total RNA was reverse transcribed and quantified using a LightCycler 480. TATAA box protein was used as an internal control for normalization. Each sample was run in triplicate. Normalized gene expression values were obtained using LightCycler 480 software version 1.5.1, using the basic relative quantification method.

4.3 Statistics

All statistical analyses consisted of testing the statistical significance of a result and visualizing a summary of the results. Plots for the data analyzed with the Student's *t*-test were made with Microsoft Excel. Plots for all other data were made with the facilities provided by the R environment for statistical computing.

4.3.1 Student's *t*-test (I, II)

To test the statistical significance of normally distributed values in exactly two groups, a non-paired two-tailed Student's *t*-test was performed. The data was visualized using bar charts, with the height of the bar representing the mean and the error bars representing the standard error of the means.

4.3.2 Mann-Whitney-Wilcoxon rank sum test (II)

Continuous distributions of values in two groups that could not be assumed to be normally distributed were tested for statistical significance using the Mann-Whitney-Wilcoxon rank sum test. The data was visualized in box-and-whiskers plots, with the bottom and the top of the box representing the first and third quartiles and the band inside the box representing the median. The lower and upper whisker represent the lowest datum within 1.5× the interquartile range of the lower quartile, and the highest datum within 1.5× the interquartile range of the upper quartile. Data points outside of those ranges were not visualized.

4.3.3 Chi-square test (II, III)

Correlation between categorical data was tested for statistical significance with the χ^2-test. In the case of two variables with two possible values each, the 2×2 contingency table was visualized as a bar plot, where the height for each bar represents the number of observations in each table cell.

4.3.4 Log-rank test (III)

Censured survival data was tested for statistical significance using the log-rank test. All *p* values are two-sided. Survival curves were plotted using the R `survival` package.

4.3.5 Pearson's correlation (III)

To estimate the correlation between two continuous variables without assuming any distribution, the Pearson's product-moment correlation coefficient was used. Such data was visualized as a scatter plot.

4.4 Computational methods

4.4.1 Image quantification (I, II)

All manual image quantifications – microscopy images, immunoblots, thin layer chromatography plates – were performed using the standard distribution of ImageJ, version 1.48 or later, without using any additional plugins (Schneider et al. 2012).

4.4.2 Data analysis (I, II, III)

For the Student's *t*-test, either Microsoft Excel or R was used. For all other statistical tests, the built-in R facilities were used. All data visualizations in Original Publications I and II were done with Microsoft Excel. All data visualizations in Original Publication III were done with R. High-throughput data was pre-processed with standard command line utilities. The data base engine used was SQLite.

4.4.3 Reproducible computation (III)

The tool for reproducible computation Lir was implemented in Bash using the command line tools available in the distribution of Noweb, as well as the standard command line programs GNU-Sed and GNU-Awk. SWI-Prolog was used for parsing the pipeline representation emitted by Noweb and generating HTML. Pandoc was used for generating the final human-readable document.

5 Results and Discussion

In each subsection of "Results and Discussion", the original publications containing all data are referred to with I, II, and III, as numbered in "Original Publications" on page vi. All following figures are unpublished data related to publication II.

5.1 NDRG1 regulates LDLR trafficking (I)

The protein N-myc downstream-regulated gene 1 (NDRG1) is widely expressed, particularly in the nervous system. Subcellularly, it is mostly localized in the cytoplasm and the endosomes. NDRG1 was first identified as a potential regulator of cellular cholesterol levels in studies of the neurodegenerative disease Niemann-Pick Type C. Reddy et al. (2006) found that in fibroblasts from patients suffering from this lysosomal storage disease NDRG1 mRNA was elevated. Furthermore, in an Affimetrix gene expression profiling of NPC1-deficient murine macrophages by Hölttä-Vuori et al. (2012), NDRG1 was identified as a gene highly induced by cholesterol loading (Elina Ikonen, unpublished observation). Finally, an RNAi screen by Bartz et al. (2009) studying cellular lipid phenotypes showed that 1,1′-dioctadecyl-3,3,3′,3′-tetramethyl-indocarbocyanine-perchlorate – labeled LDL (DiI-LDL) uptake by HeLa cells was stimulated by the knockdown of *NDRG1*. Together, these findings suggested a connection between NDRG1 and cellular cholesterol homeostasis. We studied the role of NDRG1 in lipid homeostasis by testing the effects of altering NDRG1 protein levels on cellular cholesterol, low-density lipoprotein (LDL) uptake, and low-density lipoprotein receptor (LDLR) trafficking and degradation.

5.1.1 NDRG1 alters cellular cholesterol balance

First, we tested whether NDRG1 protein levels are influenced by cellular cholesterol levels. In A431 cells, drug-induced lysosomal cholesterol accumulation resulted in a three-fold increase of total NDRG1 protein amounts. Loading the cells with cholesterol from LDL resulted in a two-fold increase of total NDRG1 protein amounts.

As NDRG1 protein levels were upregulated by an increase in lysosomal cholesterol content or total cellular cholesterol, we next tested whether the protein is needed for cholesterol homeostasis. Efficient knockdown of NDRG1 protein levels in A431 cells led to a redistribution of cholesterol from the plasma membrane (PM) to perinuclear, punctate structures. The total level of free cholesterol in the cells was largely unchanged. However, cholesterol ester levels were reduced dramatically. Since there were no major differences in cholesterol biosynthesis or efflux, we studied whether this imbalance upon NDRG1 knockdown is mediated by cholesterol uptake.

5.1.2 Diminished LDL uptake in NDRG1-depleted cells

Knockdown of NDRG1 in A431 cells resulted in an increase of total LDL receptor protein levels. Still, there was a decrease in the binding and internalization of DiI-LDL. A biotinylation of the PM LDLR showed a clear decrease of the fraction of LDLR on the

PM upon NDRG1 knockdown. Biotinylation followed by the uptake and degradation of the biotinylated LDLR showed that in NDRG1 depleted cells, LDLR is internalized and degraded at a lower rate. These results demonstrated that NDRG1 is required for the localization of LDLR to the PM and the rate at which LDL is internalized and degraded.

5.1.3 In NDRG1-depleted cells LDLR accumulates in the early endosomes

As the total levels of LDLR were increased, we tested whether LDLR is redistributed within the cell upon NDRG1 silencing. In control A431 cells LDLR is found mostly in small, punctate early endosomes and to some extent in late endosomes (LEs) and lysosomes. In NDRG1-silenced cells enlarged endosomes and lysosomes were scattered in the cytoplasm; LDLR was sequestered in the abnormal early endosomes with a high free cholesterol content. Silencing prenylated Rab acceptor protein 1 (PRA1), an interaction partner for NDRG1, showed a similar phenotype. In NDRG1-depleted cells, PRA1 protein levels were also down; overexpressing PRA1 in NDRG1-depleted cells could partially rescue the LDLR phenotype.

These findings showed the abnormal sequestration of LDLR upon NDRG1 knockdown or knockdown of its interaction partner PRA1. LDLR accumulated in abnormal early endosome antigen 1 (EEA1)-positive endosomal vesicles.

5.1.4 NDRG1 is required for endosomal maturation

NDRG1-silenced cells contained enlarged multivesicular bodies with more intraluminal vesicles, as observed by electron microscopy. In parallel, total ceramide levels in these cells were increased. Fluorescently-labeled ceramide incorporation showed that in NDRG1-depleted cells, ceramide was enriched in enlarged multi-vesicular bodies (MVBs), and was retained there longer. Feeding A431 cells with radioactively labeled sphingomyelin revealed increased levels of sphyngomyelin-derived ceramide in NDRG1-depleted cells.

These results suggested that NDRG1 function is required for proper endosomal maturation. Despite the reduction of ESCRT components, NDRG1 deficiency increases the number of intraluminal vesicles in MVBs. These vesicles are ceramide enriched: this suggests that this endosomal ceramide accumulation triggers the formation of intraluminal vesicles.

5.1.5 LDLR recycling and degradation depends on NDRG1

NDRG1 silencing in A431 cells enhanced the colocalization of LDLR with ubiquitin in enlarged endosomes. In HEK293T cells, co-transfection with NDRG1 siRNA and HA-tagged LDLR plasmid showed that NDRG1 depletion increased the amounts of ubiquitynated LDLR. In A431 cells, preventing LDLR ubiquitination by silencing inducible degrader of the LDL-receptor (IDOL) increased both PM and total LDLR levels, as expected. In NDRG1-depleted A431 cells, IDOL depletion reversed the decrease in LDLR on the PM and alleviated the endosomal sequestration of LDLR; it also rescued, to some degree, the levels of cholesterol esters.

These results demonstrate that LDLR is ubiquitynated in NDRG1-depleted cells and accumulates in enlarged endosomes and MVBs. When LDLR ubiquitination is prevented, the receptor is not abnormally sequestered and can recycle to the PM and cholesterol uptake is rescued.

As epidermal growth factor receptor (EGFR) has to be transported to the lysosomes to be degraded, defects in the pathway to the lysosome could influence its degradation. A recent finding by Kovacevic et al. (2016) suggests one example of a possible action of NDRG1 through this mechanism. In pancreatic cancer cells and colon cancer cells, overexpression of NDRG1 decreased the expression and activation of EGFR, HER2, and HER3 in response to EGF. Furthermore, it inhibited the formation of EGFR/HER2 and HER2/HER3 heterodimers, and decreased the activation of downstream MAPKK in response to EGF. Anti-tumor agents that up-regulate NDRG1 inhibited EGFR, HER2, and HER3 in cancer cells.

5.1.6 Reduced LDL uptake and Olig2 amounts in NDRG1-depleted oligodendrocytes

Because of the abundant expression of NDRG1 in the nervous system, we depleted NDRG1 in oligodendrocytic cells. Silencing of NDRG1 in murine Oli-neu cells reduced DiI-LDL binding and uptake. Conversely, silencing IDOL resulted in increase of DiI-LDL binding and uptake. Moreover, depletion of IDOL restored DiI-LDL binding and uptake in NDRG1-silenced cells. The number of outgrowths characteristic of oligodendrocytes were diminished in NDRG1-depleted Oli-neu cells. The levels of Olig2, important regulator of oligodendrocyte differentiation and myelination, were significantly reduced. This effect was reversed upon IDOL depletion in NDRG1-depleted Oli-neu cells.

These findings suggest that NDRG1 is important for oligodendrocyte cholesterol balance and differentiation. This effect could be mediated through IDOL-sensitive LDLR family members VLDLR and ApoE receptor 2, both sharing a strong homology with the LDLR (Herz and Bock 2002). Hong et al. (2010) have reported that IDOL is involved in the ubiquitination and degradation of both receptors.

5.2 StARD3 and cholesterol in breast cancer (II)

The *STARD3* gene is located in the genomic region of *ERBB2* in chromosome 17 (17q12-q21) and is frequently amplified in ErbB2-positive cancers (Kauraniemi and Kallioniemi 2006; Staaf et al. 2010). The RNAi studies by Kao and Pollack (2006) and Sahlberg et al. (2013) suggested that StARD3 contributes to the proliferation of ErbB2-positive cancer cells. We found that ErbB2-positive breast cancer cell lines KPL-4 and BT-474 co-overexpress StARD3 (fig. 8). To study the mechanisms by which StARD3 in particular contributes to cancer progression, we needed a system in which we can decouple it from ErbB2. This was done either by knocking down StARD3 in an overexpressing cell line, or overexpressing it in a ErbB2-negative cell line.

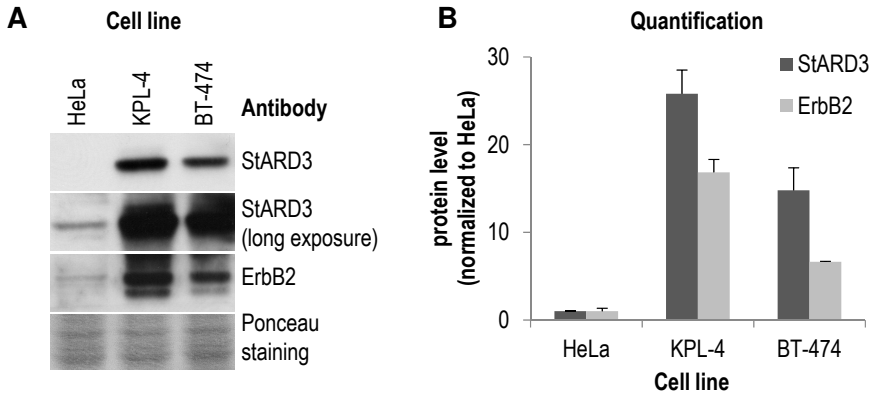


Figure 8: Comparison of StARD3 and ErbB2 protein levels in cervical cancer cell line HeLa and ErbB2-positive breast cancer cell lines KPL-4 and BT-474. **A**: Samples from cells grown in normal culture conditions from each of the cell lines were collected and the same amount of protein was separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The filter was stained with Ponceau dye after transfer. The filter was cut horizontally and the pieces immunoblotted against StARD3 and ErbB2 antibodies, respectively. **B**: The quantification shows the StARD3 and ErbB2 cellular protein levels.

5.2.1 ErbB2-positive breast cancer cell lines require StARD3 overexpression for survival

Transient knockdown of StARD3 and ErbB2 in BT-474 cells had a cumulative negative effect on cell proliferation. In BT-474 cells, 3 d of siRNA knockdown of either StARD3 or ErbB2 decreased cell amounts. (fig. 9, A and B). To create a cell line with stable knockdown of either StARD3 or ErbB2, we transduced KPL-4 cells with shRNA lentiviral vectors. Lentiviral transduction of KPL-4 cells with shRNA resulted in efficient knockdown of StARD3 and ErbB2 protein levels. The knockdown of StARD3 and ErbB2 levels had a strong negative effect on cell growth (fig. 9, C and D). This efficient knockdown inhibited cell proliferation so strongly that the generation of a stable cell line was not possible. Within 8 d to 12 d of transfection all cells died.

In comparison, the results by Kishida et al. (2004) show that StARD3 was largely dispensable in healthy tissue. In mice homozygous for a StARD3 mutant allele, there was no clear phenotype, only non-lethal defects in sterol trafficking. However, our result strongly suggested that the overexpression of StARD3 is required by ErbB2-overexpressing breast cancer cells.

5.2.2 Stable overexpression of StARD3 alters cell morphology

The Her2-negative breast cancer cell line, MCF-7, was used to generate a stable clone overexpressing StARD3-GFP. As a control, a cell line overexpressing GFP was generated. The StARD3-GFP exhibited a punctate distribution and partly colocalized with the late-

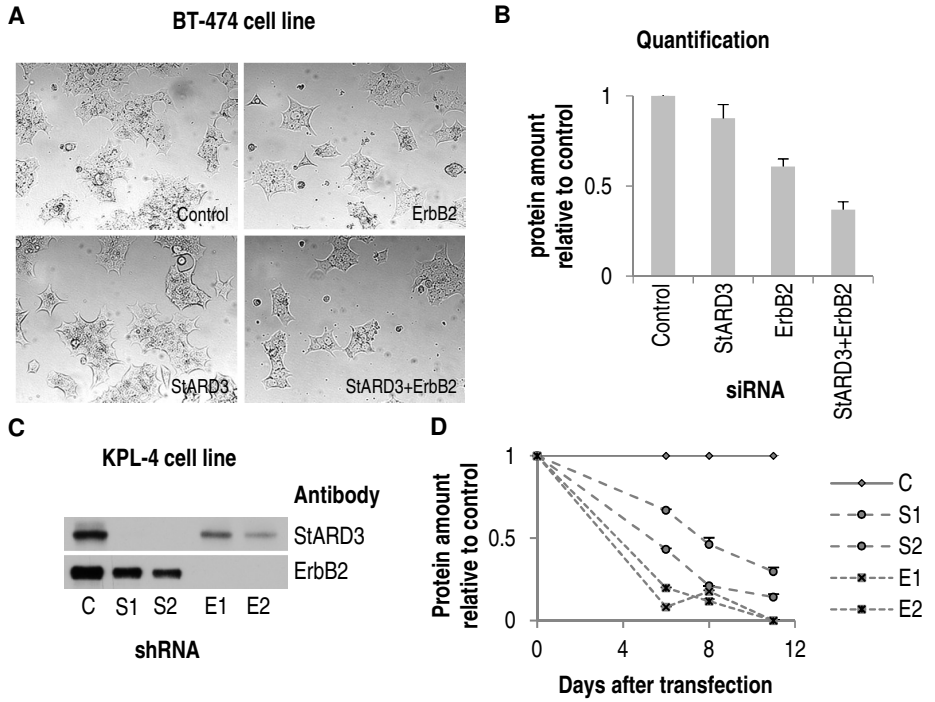


Figure 9: Knockdown of StARD3 and ErbB2 protein levels in Erb2-positive breast cancer cell lines BT-474 and KPL-4. **A**: Differential interference contrast (DIC) images of live BT-474 cells on plastic after 3 d of siRNA knockdown of either StARD3, ErbB2, or both. **B**: BT-474 cells treated as in (A) and grown on 12-well plates were lyzed in equal amounts of lysis buffer. Total protein amount in each sample was used as measure of cell amounts. $n = 5$ for all samples but StARD3, where $n = 3$. For all conditions compared to control, $p < 0.05$. **C**: Samples of KPL-4 cells after 6 d of lentiviral transduction with shRNAs were separated by SDS-PAGE and immunoblotted using StARD3 and ErbB2 antibodies. "C" stands for control shRNA, "S1" and "S2" are two different shRNAs targeting StARD3, and "E1" and "E2" are two different shRNAs targeting ErbB2. **D**: Relative proliferation of KPL-4 cells after shRNA knockdown of StARD3 and ErbB2. Cells were collected in lysis buffer 6, 8 and 11 d after lentiviral transduction and total protein was measured.

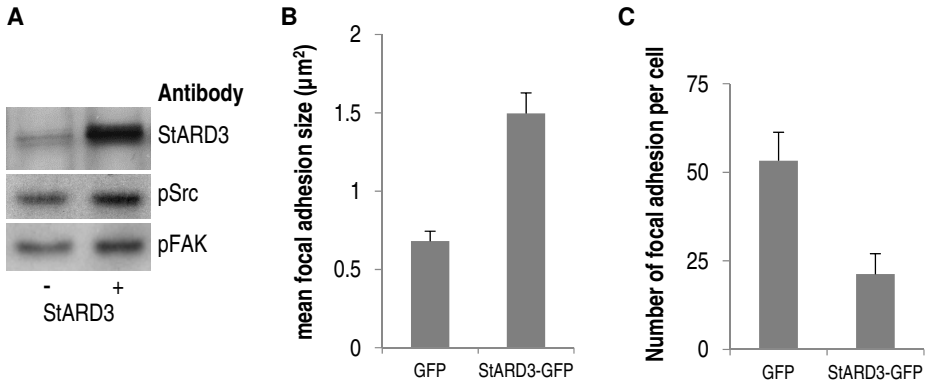


Figure 10: Effects of StARD3 overexpression on focal adhesions in HeLa cells. **A**: Immunoblots of activated Src (Y416) and FAK (Y397) in HeLa cells after 3 d of GFP or StARD3-GFP transient overexpression. Cells were trypsinized and allowed to readhere for 50 min on fibronectin-coated plastic. **B**: Quantification of the size of focal adhesions in HeLa cells. Cells were treated as in (A) and plated on fibronectin-coated coverslips. Representative images in section 5.2.2. **C**: Quantification of the number of focal adhesions per cell in HeLa cells as in (B). For (B) and (C), $n = 33$ cells for GFP and $n = 38$ cells for StARD3-GFP.

endosomal marker Lamp-1. Some of the StARD3-GFP localized to the PM, and confocal microscopy revealed a punctate pattern that could indicate PM domains and small vesicles immediately below the PM. Phase-contrast and green fluorescence images showed that the StARD3-GFP expressing cells had a strikingly different morphology compared to the parental MCF-7 cell line or the GFP overexpressing controls. They grew in clusters of rounded-up cells that did not spread on the substratum and lacked contact inhibition: a malignant cell growth pattern.

The altered adhesion of MCF-7 cells was in line with the effects of transient StARD3 overexpression on cell morphology and focal adhesions in HeLa cells. In HeLa cells, transient StARD3 overexpression activated both proto-oncogene tyrosine-protein kinase Src (Src) and focal adhesion kinase (FAK), as indicated by the increased the phosphorylation of Src on tyrosine 416 and FAK on tyrosine 397. (fig. 10). The StAR-related lipid transfer protein 3 (StARD3) overexpressing cells were also less spread out, with fewer, but larger irregularly shaped focal adhesions (fig. 10 and section 5.2.2).

5.2.3 StARD3 overexpression alters cholesterol balance

The StARD3-GFP overexpressing MCF-7 cell clusters showed increased filipin staining intensity, especially on the PM. However, the total amount of cellular free cholesterol was similar to that of control cells, suggesting that the cholesterol distribution was altered. In StARD3-GFP overexpressing cells, the levels of the rate-limiting enzyme of cholesterol synthesis HMGR was increased. In addition, both the precursor and the mature form of SREBP2, the key transcriptional regulator of cholesterol metabolism, were increased. Both

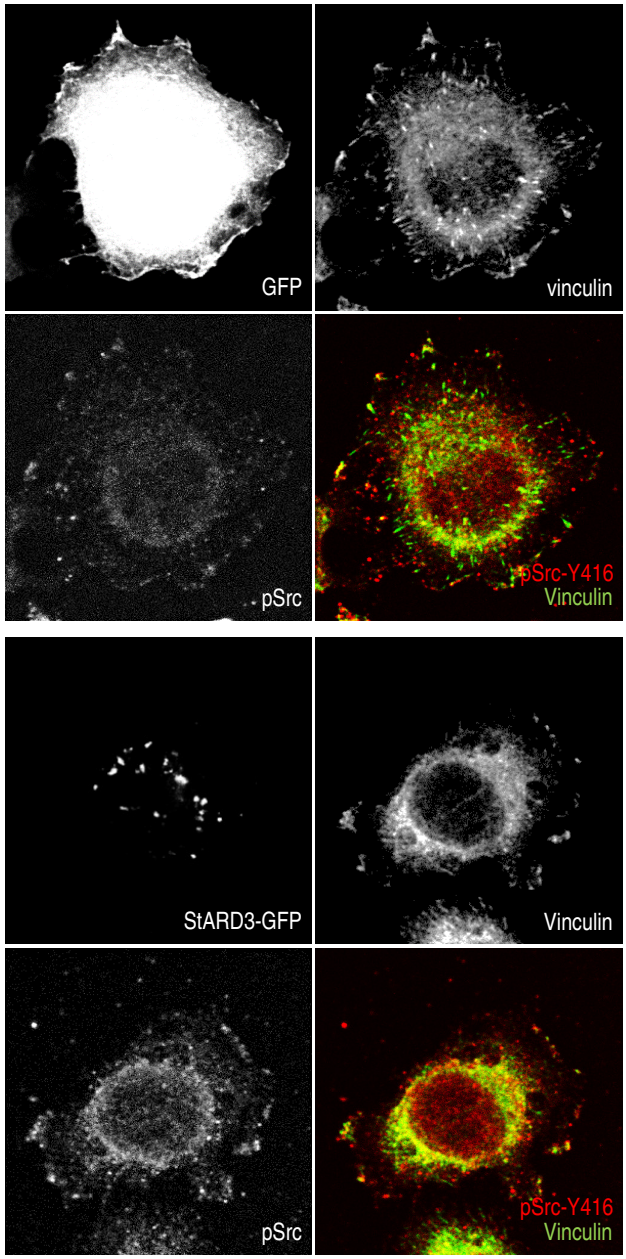


Figure 11: HeLa cells overexpressing StARD3-GFP are rounded up and have abnormal focal adhesions. The cells were transiently transfected with soluble GFP (top) or StARD3-GFP (bottom) for 3 d. Cells were trypsinized and let readhere for 50 min on fibronectin-coated coverslips. Immunofluorescent staining for vinculin was with A568, and for pSrc with a Cy5 fluorescently-labeled secondary antibodies. Confocal images were acquired with a Leica TCS SP2 AOBS confocal microscope with a 63× oil immersion objective.

findings suggest that the endoplasmic reticulum (ER) was cholesterol depleted as a result of StARD3 overexpression. Moreover, streptolysin O (SLO) resistance of StARD3-GFP overexpressing cells was decreased, suggesting an increase of cholesterol in the PM.

5.2.4 Signaling in StARD3 overexpressing cells does not depend on growth stimuli

Within 5 h of removal of serum and insulin from the cell culture medium, StARD3-GFP overexpressing cells were found to revert to a flat, adherent morphology reminiscent of that of the control cells. The GFP-overexpressing control cells did not show a change in morphology within this time frame. In complete medium, the rounded StARD3-GFP overexpressing cells had fewer, larger focal adhesions. StARD3-GFP overexpressing cells under serum and insulin deprivation conditions flattened and adhered to the substratum, developing larger, more prominent, irregularly shaped focal adhesions, as observed by confocal microscopy. In full serum conditions, there was no marked difference between the total levels or the phosphorylation status of focal adhesion kinase. Under serum and insulin deprivation, StARD3 overexpressing cells had significantly higher focal adhesion kinase total protein levels and phosphorylation than control cells.

5.2.5 StARD3-overxpressing cells maintain increased cellular and PM cholesterol in the absence of serum and insulin

Upon removal of serum and insulin from the medium for 5 h and 24 h, cells with high levels of StARD3-GFP showed a robust increase in filipin intensity compared to control GFP cells or cells with low levels of StARD3-GFP expression. The filipin intensity of control cells decreased from 5 h to 24 h. The filipin intensity of StARD3-GFP cells however did not decrease. Under these conditions, StARD3-GFP cells had higher levels of cholesterol than control cells, as determined by a biochemical lipid determination and SLO resistance. This difference was abolished when lovastatin, an inhibitor of HMGR, was added to the cells, indicating that StARD3-overexpressing cells exhibit increased *de novo* cholesterol synthesis in the absence of growth factors.

5.2.6 High StARD3 protein levels in breast tumor samples

We generated a StARD3 polyclonal antibody, and cell culture experiments showed that it is specific. It was also tested in patient tissue samples. In a cohort of 1325 Finnish breast cancer patients (Joensuu et al. 2003), high StARD3 immunoreactivity in tumor samples associated with ErbB2 immunoreactivity and *HER2* amplification, as well as positive Src phosphorylation. High StARD3 immunoreactivity also associated with factors related to poor disease outcome: high proliferation rate, bigger tumor size, nodal metastases at the time of diagnosis, estrogen and progesterone receptor negativity and p53 protein expression. In an independent breast cancer cohort of 895 patients (Joensuu et al. 2006), these results were reproduced. For a subset of these samples, RNA was available. A qPCR analysis of HMGR mRNA levels showed that the highest transcript levels were found in

the highly StARD3-positive tumors. High StARD3 protein level was strongly associated with decreased cumulative patient survival in both cohorts. Thus, our findings in MCF-7 cells were strengthened by the findings in patient breast cancer tumor samples from two independent patient cohorts.

5.3 TCGA breast cancer database analysis (III)

The results so far suggested a role for NDRG1 in cancer progression, possibly through the effects that abnormal lysosomal formation have on ErbB-family receptor recycling (Tan et al. 2015a,b). Our results also demonstrated that StARD3 overexpression in ErbB2-positive breast cancer plays a role in breast cancer progression. Since the TCGA database contains comprehensive molecular portraits of breast cancer tumors and clinical patient data, we analyzed the available data searching for further clues to the role of these proteins in cancer progression. We also included in the computational analysis *lysosomal-associated transmembrane protein 4B (LAPTM4B)*, because it regulates ceramide-dependent cell death pathways in cancer cells (Blom et al. 2015; Trajkovic et al. 2008).

5.3.1 *STARD3*, *NDRG1*, and *LAPTM4B* in cancer progression

The strong correlation between *STARD3* and *ERBB2* amplification and expression was reproduced in the TCGA breast cancer dataset. This finding provided us with a validation of the analysis methods that we aimed to use in the study. We found that *NDRG1* and *LAPTM4B* were co-amplified, and there was a positive correlation between the mRNA levels of the two genes. *LAPTM4B* amplification and high *NDRG1* mRNA levels both had a statistically significant negative effect on breast cancer patient survival. This correlation between the two genes is not too surprising: they are on the same chromosome arm and it is likely that they are coamplified in some cancer patients.

Given the role of LAPTM4B in ceramide export from LEs reported by Blom et al. (2015), this finding reinforces the hypothesis that altered ceramide trafficking is the common denominator upon deregulation of *NDRG1* or *LAPTM4B* expression. However, in this dataset, co-overexpression of *NDRG1* and *LAPTM4B* did not exhibit an additive negative effect on patient survival.

5.3.2 Reproducible computational data analysis

In the course of this data mining study, our attempts to document all computational analyses and ensure their reproducibility were hampered. This was caused by a lack of software for documenting and automating a computational analysis that combines different tools. We therefore proposed a method and provided Lir, a software tool for reproducible computing. We demonstrated that a computational data analysis can be documented, shared, and reproduced without increasing the work load. Instead, by allowing a combination of computational data analysis tools to be used in the same coherent workflow, we were able to minimize the total amount of work to obtain biologically relevant results.

The positive effect of documenting and automating a heterogeneous workflow in a literate programming style is synergistic. By automating the workflow, we have already defined it formally and thus provided the most basic documentation. An automated workflow requires no manual effort to run, encouraging an iterative, exploratory approach. An automated workflow also requires no manual effort to re-run, increasing its reproducibility. Since it is easy to share the workflow in full, other scientists can validate and reproduce the results, or adapt the analysis for their data. Finally, a workflow documented in a literate programming style can follow the established structure of an experimental protocol: Aims; Experimental design; Methods; Results; Discussion. The clear structure and its familiarity help communicate the goals and results of the analysis to other scientists.

6 Conclusions

In this work we studied the role of proteins that have been implicated in cancer progression and in endocytic lipid trafficking, and how these two processes may be functionally connected. Two possible mechanisms emerged. On the one hand, through regulation of multi-vesicular body (MVB) formation, N-myc downstream-regulated gene 1 (NDRG1) functions as a regulator of cell surface receptor recycling and degradation. On the other hand, StAR-related lipid transfer protein 3 (StARD3) overexpression promotes cell proliferation by deregulating the cholesterol content of the late endocytic membrane. While the two mechanisms do not seem to be tightly coupled, they both further the understanding of endocytic trafficking. An analysis of genome wide breast cancer patient data provided clues to other understudied proteins involved in cancer progression and endocytic trafficking. The data analysis was presented, published, and made publicly available: it can serve as a starting point to further analysis of the same dataset, or as a template for analysing another dataset with the same methods.

In particular, we identified NDRG1 as a regulator of MVB formation and endosomal low-density lipoprotein receptor (LDLR) trafficking. Its function in endosomal dynamics suggested that it could be involved in tumorigenesis. Indeed, the results by Kovacevic et al. (2016) showed that NDRG1 overexpression decreases the expression and activation of ErbB-family receptors in response to EGF stimulation, inhibits the formation of ErbB-family receptor heterodimers, and decreases the activation of downstream MAPK kinase in response to EGF. These effects might be partly explained with the positive correlation between NDRG1 and the endosomal ceramide regulator lysosomal-associated transmembrane protein 4B (LAPTM4B) that we found in the TCGA breast cancer tumor dataset. It is possible that the LAPTM4B-mediated inhibition of EGFR degradation offsets the negative effects of NDRG1 upregulation on EGF signaling.

We established that elevated levels of StARD3 protein even in the absence of ErbB2 overexpression alter the cholesterol balance of breast cancer cells, enhance oncogenic signaling, and cause a phenotype characteristic of malignant cells. Importantly, the enhanced oncogenic signaling we observed upon StARD3 overexpression was independent of external growth stimuli, further reinforcing the conclusion that StARD3 overexpression in tumors is a necessary component of ErbB2-positive cancer progression. The recent results by Wilhelm et al. (2016) and Wilhelm et al. (2017) showing that StARD3 tethers ER to late endosomes provide a possible mechanistic explanation for the uncoupling of external growth stimuli and ER functions such as cholesterol synthesis that we observed in StARD3-overexpressing cells. If these ER-endosome contacts help deplete the ER membrane of cholesterol, this would activate the sterol regulatory element-binding protein (SREBP) feedback mechanism and enhance *de novo* cholesterol synthesis. There is in fact recent evidence for another cholesterol binding protein, ORP1L, in facilitating such cholesterol transfer (Eden et al. 2016).

Finally, we demonstrated that automating, documenting, and presenting a data analysis not only can improve reproducibility, but also can minimize work, effort, and room for potential errors. Of course, the positive effects we claim would have to be empirically measured; however, reliable methods for empirical studies of the process of writing

computer programs are still to be developed and applied systematically. An interesting question is whether traditional, "wet lab" experimental methods and protocols would benefit from a formal description language that can be automatically parsed by machines. The goal would not be to automate the carrying out of experiments. It would be already very advantageous if the protocols and the results of all experiments within a lab can be searched and summarized, maybe even shared with collaborators and the wider scientific community.

Acknowledgements

All work for this dissertation was conducted at the Intracellular Cholesterol Trafficking lab of prof. Elina Ikonen. The lab has a double affiliation with the University of Helsinki, Faculty of Medicine, Department of Anatomy, and with the Minerva Foundation Institute for Medical Research, both located at Biomedicum Helsinki.

The work was in part financed by the Doctoral Programme in Biomedicine, formerly the Helsinki Biomedical Graduate School; as well as grants from the Ida Montinin Säätiö, K. Albin Johanssons stiftelse, Paulon Säätiö, Biomedicum Helsinki Foundation, and The Finnish Society of Sciences and Letters. The preparation of this dissertation was financed by a grant by the Doctoral Programme in Biomedicine, and a grant by the Minerva Foundation Institute for Medical Research.

I would like to thank all collaborators that have made this work possible, in particular prof. Heikki Joensuu, Harri Sihto, and prof. Sampsa Hautaniemi.

I would like to thank the thesis pre-examiners, Petri Auvinen and Simon Anders, for taking the time and the effort to read the thesis and provide invaluable comments.

I would like to thank my thesis supervisor, prof. Elina Ikonen, who has contributed to this work in every possible way: from agreeing to supervise the thesis, to providing the environment necessary for academic research, to providing education, guidance, and support.

I would like to thank Anna Uro and Pipsa Kaipainen for all the experiments they have performed, and Anna Uro in particular for teaching me so many things about our lab in particular and experimental work in general.

I would like to thank all current and former members of the lab for the help, the patience, and the good times. In particular, Maurice Jansen, my hands-on mentor; Maarit Hölttä-Vuori for her invaluable advice; and Tomas Blom for sacrificing a considerable amount of his time to talk and listen. The list goes on.

I would like to thank my family for putting up with me and helping me during this time: Lotta, Alli, Aatos, Maja, and Toma. I would like to thank my parents, Bojko and Boyana, for never pushing and always believing; and my grandparents, Anna, Boris, Liljana, and Pavel, for the important role they have had in my upbringing and my education.

References

Alpy, F. and Tomasetto, C. (2006). "MLN64 and MENTHO, two mediators of endosomal cholesterol transport." *Biochemical Society Transactions* 34.Pt 3, pp. 343–345.

Alpy, F., Latchumanan, V. K., Kedinger, V., Janoshazi, A., Thiele, C., Wendling, C., Rio, M.-C., and Tomasetto, C. (2005). "Functional characterization of the MENTAL domain." *Journal of Biological Chemistry* 280.18, pp. 17945–17952.

Bandyopadhyay, S., Pai, S. K., Gross, S. C., Hirota, S., Hosobe, S., Miura, K., Saito, K., Commes, T., Hayashi, S., Watabe, M., and Watabe, K. (2003). "The Drg-1 gene suppresses tumor metastasis in prostate cancer." *Cancer Research* 63.8, pp. 1731–1736.

Bandyopadhyay, S., Pai, S. K., Hirota, S., Hosobe, S., Takano, Y., Saito, K., Piquemal, D., Commes, T., Watabe, M., Gross, S. C., Wang, Y., Ran, S., and Watabe, K. (2004). "Role of the putative tumor metastasis suppressor gene Drg-1 in breast cancer progression." *Oncogene* 23.33, pp. 5675–5681.

Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., Wirkner, U., Erfle, H., Muckenthaler, M., Pepperkok, R., and Runz, H. (2009). "Identification of cholesterol-regulating genes by targeted RNAi screening." *Cell Metabolism* 10.1, pp. 63–75.

Bhakdi, S., Tranum-Jensen, J., and Sziegoleit, A. (1985). "Mechanism of Membrane Damage by Streptolysin-O". *Infection and Immunity* 47.1, pp. 52–60.

Bligh, E. G. and Dyer, W. J. (1959). "A Rapid Method of Total Lipid Extraction and Purification". *Canadian Journal of Biochemistry and Physiology* 37.8. PMID: 13671378, pp. 911–917.

Bloch, K. (1965). "The biological synthesis of cholesterol". *Science* 150.3692, pp. 19–28.

Blom, T., Li, S., Dichlberger, A., Bäck, N., Kim, Y. A., Loizides-Mangold, U., Riezman, H., Bittman, R., and Ikonen, E. (2015). "LAPTM4B facilitates late endosomal ceramide export to control cell death pathways." *Nature Chemical Biology* 11.10, pp. 799–806.

Börnigen, D., Moon, Y. S., Rahnavard, G., Waldron, L., McIver, L., Shafquat, A., Franzosa, E. A., Miropolsky, L., Sweeney, C., Morgan, X. C., Garrett, W. S., and Huttenhower, C. (2015). "A reproducible approach to high-throughput biological data acquisition and integration." *PeerJ* 3, e791.

Brown, D. N., Caffa, I., Cirmena, G., Piras, D., Garuti, A., Gallo, M., Alberti, S., Nencioni, A., Ballestrero, A., and Zoppoli, G. (2016). "Squalene epoxidase is a bona fide oncogene by amplification with clinical relevance in breast cancer." *Scientific Reports* 6, p. 19435.

Carstea, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M. A., Pavan, W. J., Krizman, D. B., Nagle, J., Polymeropoulos, M. H., Sturley, S. L., Ioannou, Y. A., Higgins, M. E., Comly, M., Cooney, A., Brown, A., Kaneski, C. R., Blanchette-Mackie, E. J., Dwyer, N. K., Neufeld, E. B., Chang, T. Y., Liscum, L., Strauss 3rd, J., Ohno, K., Zeigler, M., Carmi, R., Sokol, J., Markie, D., O'Neill, R. R., van Diggelen, O. P., Elleder, M., Patterson, M. C., Brady, R. O., Vanier, M. T., Pentchev, P. G., and Tagle, D. A. (1997). "Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis." *Science* 277.5323, pp. 228–231.

Charman, M., Kennedy, B. E., Osborne, N., and Karten, B. (2010). "MLN64 mediates egress of cholesterol from endosomes to mitochondria in the absence of functional Niemann-Pick Type C1 protein." *Journal of Lipid Research* 51.5, pp. 1023–1034.

Dijkstra, E. W. (1989). "On the cruelty of really teaching computing science". *Communications of the ACM* 32.12, pp. 1398–1404.

Eden, E. R., Sanchez-Heras, E., Tsapara, A., Sobota, A., Levine, T. P., and Futter, C. E. (2016). "Annexin A1 Tethers Membrane Contact Sites that Mediate ER to Endosome Cholesterol Transport". *Developmental Cell* 37.5, pp. 473–483.

Freed-Pastor, W. A., Mizuno, H., Zhao, X., Langerød, A., Moon, S.-H., Rodriguez-Barrueco, R., Barsotti, A., Chicas, A., Li, W., Polotskaia, A., Bissell, M. J., Osborne, T. F., Tian, B., Lowe, S. W., Silva, J. M., Børresen-Dale, A.-L., Levine, A. J., Bargonetti, J., and Prives, C. (2012). "Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway." *Cell* 148.1-2, pp. 244–258.

Goldstein, J. L. and Brown, M. S. (2009). "The LDL receptor." *Arteriosclerosis, Thrombosis, and Vascular Biology* 29.4, pp. 431–438.

Grieb, B. C., Chen, X., and Eischen, C. M. (2014). "MTBP Is Overexpressed in Triple-Negative Breast Cancer and Contributes to Its Growth and Survival". *Molecular Cancer Research* 12.9, pp. 1216–1224.

Guan, R. J., Ford, H. L., Fu, Y., Li, Y., Shaw, L. M., and Pardee, A. B. (2000). "Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer." *Cancer Research* 60.3, pp. 749–755.

Haskins, J. W., Zhang, S., Means, R. E., Kelleher, J. K., Cline, G. W., Canfrán-Duque, A., Suárez, Y., and Stern, D. F. (2015). "Neuregulin-activated ERBB4 induces the SREBP-2 cholesterol biosynthetic pathway and increases low-density lipoprotein uptake." *Science Signaling* 8.401, ra111.

Heino, S., Lusa, S., Somerharju, P., Ehnholm, C., Olkkonen, V. M., and Ikonen, E. (2000). "Dissecting the role of the golgi complex and lipid rafts in biosynthetic transport of cholesterol to the cell surface." *Proceedings of the National academy of Sciences of the United States of America* 97.15, pp. 8375–8380.

Herz, J. and Bock, H. H. (2002). "Lipoprotein Receptors in the Nervous System". *Annual Review of Biochemistry* 71.1, pp. 405–434.

Hölttä-Vuori, M., Vainio, S., Kauppi, M., Van Eck, M., Jokitalo, E., and Ikonen, E. (2012). "Endosomal Actin Remodeling by Coronin-1A Controls Lipoprotein Uptake and Degradation in Macrophages". *Circulation Research* 110.3, pp. 450–455.

Hölttä-Vuori, M., Alpy, F., Tanhuanpää, K., Jokitalo, E., Mutka, A.-L., and Ikonen, E. (2005). "MLN64 is involved in actin-mediated dynamics of late endocytic organelles." *Molecular Biology of the Cell* 16.8, pp. 3873–3886.

Hölttä-Vuori, M., Sezgin, E., Eggeling, C., and Ikonen, E. (2016). "Use of BODIPY-Cholesterol (TF-Chol) for Visualizing Lysosomal Cholesterol Accumulation". *Traffic* 17.9, pp. 1054–1057.

Hölttä-Vuori, M., Uronen, R.-L., Repakova, J., Salonen, E., Vattulainen, I., Panula, P., Li, Z., Bittman, R., and Ikonen, E. (2008). "BODIPY-Cholesterol: A New Tool to Visualize Sterol Trafficking in Living Cells and Organisms". *Traffic* 9.11, pp. 1839–1849.

Hong, C., Duit, S., Jalonen, P., Out, R., Scheer, L., Sorrentino, V., Boyadjian, R., Rodenburg, K. W., Foley, E., Korhonen, L., and al., et (2010). "The E3 Ubiquitin Ligase IDOL Induces the Degradation of the Low Density Lipoprotein Receptor Family Members VLDLR and ApoER2". *Journal of Biological Chemistry* 285.26, pp. 19720–19726.

Horton, J. D., Goldstein, J. L., and Brown, M. S. (2002). "SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver." *Journal of Clinical Investigation* 109.9, pp. 1125–1131.

Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D. P., Kania, R., Schaeffer, M., St Pierre, S., and al., et (2008). "Big data: The future of biocuration". *Nature* 455.7209, pp. 47–50.

Ikonen, E. (2008). "Cellular cholesterol trafficking and compartmentalization". *Nature Reviews Molecular Cell Biology* 9.2, pp. 125–138.

Jo, Y. and Debose-Boyd, R. A. (2010). "Control of cholesterol synthesis through regulated ER-associated degradation of HMG CoA reductase." *Critical Reviews in Biochemistry and Molecular Biology* 45.3, pp. 185–198.

Joensuu, H., Isola, J., Lundin, M., Salminen, T., Holli, K., Kataja, V., Pylkkänen, L., Turpeenniemi-Hujanen, T., Smitten, K. von, and Lundin, J. (2003). "Amplification of erbB2 and erbB2 Expression Are Superior to Estrogen Receptor Status As Risk Factors for Distant Recurrence in pT1N0M0 Breast Cancer: A Nationwide Population-based Study". *Clinical Cancer Research* 9.3, pp. 923–930.

Joensuu, H., Kellokumpu-Lehtinen, P.-L., Bono, P., Alanko, T., Kataja, V., Asola, R., Utriainen, T., Kokko, R., Hemminki, A., Tarkkanen, M., and al., et (2006). "Adjuvant Docetaxel or Vinorelbine with or without Trastuzumab for Breast Cancer". *New England Journal of Medicine* 354.8, pp. 809–820.

Kachhap, S. K., Faith, D., Qian, D. Z., Shabbeer, S., Galloway, N. L., Pili, R., Denmeade, S. R., DeMarzo, A. M., and Carducci, M. A. (2007). "The N-Myc down regulated Gene1 (NDRG1) Is a Rab4a effector involved in vesicular recycling of E-cadherin." *PLOS ONE* 2.9, e844.

Kanerva, K., Uronen, R.-L., Blom, T., Li, S., Bittman, R., Lappalainen, P., Peränen, J., Raposo, G., and Ikonen, E. (2013). "LDL cholesterol recycles to the plasma membrane via a Rab8a-Myosin5b-actin-dependent membrane transport route." *Developmental Cell* 27.3, pp. 249–262.

Kao, J. and Pollack, J. R. (2006). "RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes." *Genes, Chromosomes and Cancer* 45.8, pp. 761–769.

Karamouzis, M. V., Grandis, J. R., and Argiris, A. (2007). "Therapies Directed Against Epidermal Growth Factor Receptor in Aerodigestive Carcinomas". *JAMA* 298.1, p. 70.

Kauraniemi, P. and Kallioniemi, A. (2006). "Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer." *Endocr Relat Cancer* 13.1, pp. 39–49.

Keller, P. and Simons, K. (1997). "Post-Golgi biosynthetic trafficking." *Journal of cell science* 110 (Pt 24), pp. 3001–3009.

Kishida, T., Kostetskii, I., Zhang, Z., Martinez, F., Liu, P., Walkley, S. U., Dwyer, N. K., Blanchette-Mackie, E. J., Radice, G. L., and Strauss, J. F. (2004). "Targeted Mutation of the MLN64 START Domain Causes Only Modest Alterations in Cellular Sterol Metabolism". *Journal of Biological Chemistry* 279.18, pp. 19276–19285.

Kitchin, J. R. (2015). "Examples of Effective Data Sharing in Scientific Publishing". *ACS Catalysis* 5.6, pp. 3894–3899.

Knuth, D. E. (1983). *The WEB system of structured documentation*. Department of Computer Science, Stanford University.

– (1984). "Literate programming". *The Computer Journal* 27.2, pp. 97–111.

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., and al., et (2012). "Comprehensive molecular portraits of human breast tumours". *Nature* 490.7418, pp. 61–70.

Kovacevic, Z., Menezes, S. V., Sahni, S., Kalinowski, D. S., Bae, D.-H., Lane, D. J. R., and Richardson, D. R. (2016). "The Metastasis Suppressor, N-MYC Downstream-regulated Gene-1 (NDRG1), Down-regulates the ErbB Family of Receptors to Inhibit Downstream Oncogenic Signaling Pathways." *Journal of Biological Chemistry* 291.3, pp. 1029–1052.

Lachat, P., Shaw, P., Gebhard, S., van Belzen, N., Chaubert, P., and Bosman, F. T. (2002). "Expression of NDRG1, a differentiation-related gene, in human tissues." *Histochemistry and Cell Biology* 118.5, pp. 399–408.

Leisch, F. (2002). "Sweave: Dynamic generation of statistical reports using literate data analysis". In: *Compstat*. Springer, pp. 575–580.

Linder, M. D., Uronen, R.-L., Hölttä-Vuori, M., van der Sluijs, P., Peränen, J., and Ikonen, E. (2007). "Rab8-dependent recycling promotes endosomal cholesterol removal in normal and sphingolipidosis cells." *Molecular Biology of the Cell* 18.1, pp. 47–56.

Llaverias, G., Danilo, C., Mercier, I., Daumer, K., Capozza, F., Williams, T. M., Sotgia, F., Lisanti, M. P., and Frank, P. G. (2011). "Role of cholesterol in the development and progression of breast cancer." *American Journal of Pathology* 178.1, pp. 402–412.

Loftus, S. K., Morris, J. A., Carstea, E. D., Gu, J. Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M. A., Tagle, D. A., Pentchev, P. G., and Pavan, W. J. (1997). "Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene." *Science* 277.5323, pp. 232–235.

Maione, F., Oliaro-Bosso, S., Meda, C., Di Nicolantonio, F., Bussolino, F., Balliano, G., Viola, F., and Giraudo, E. (2015). "The cholesterol biosynthesis enzyme oxidosqualene cyclase is a new target to impair tumour angiogenesis and metastasis dissemination." *Scientific Reports* 5, p. 9054.

Maruyama, Y., Ono, M., Kawahara, A., Yokoyama, T., Basaki, Y., Kage, M., Aoyagi, S., Kinoshita, H., and Kuwano, M. (2006). "Tumor growth suppression in pancreatic cancer by a putative metastasis suppressor gene Cap43/NDRG1/Drg-1 through modulation of angiogenesis." *Cancer Research* 66.12, pp. 6233–6242.

Maxfield, F. R. and Wüstner, D. (2012). "Analysis of cholesterol trafficking with fluorescent probes." *Methods in Cell Biology* 108, pp. 367–393.

McMahon, H. T. and Boucrot, E. (2011). "Molecular mechanism and physiological functions of clathrin-mediated endocytosis". *Nature Reviews Molecular Cell Biology* 12.8, pp. 517–533.

Moog-Lutz, C., Tomasetto, C., Régnier, C. H., Wendling, C., Lutz, Y., Muller, D., Chenard, M. P., Basset, P., and Rio, M. C. (1997). "MLN64 exhibits homology with the steroidogenic acute regulatory protein (STAR) and is over-expressed in human breast carcinomas." *International Journal of Cancer* 71.2, pp. 183–191.

Murtola, T. J., Visvanathan, K., Artama, M., Vainio, H., and Pukkala, E. (2014). "Statin use and breast cancer survival: a nationwide cohort study from Finland." *PLOS ONE* 9.10, e110231.

Nakanishi, M., Goldstein, J. L., and Brown, M. S. (1988). "Multivalent control of 3-hydroxy-3-methylglutaryl coenzyme A reductase. Mevalonate-derived product inhibits translation of mRNA and accelerates degradation of enzyme." *Journal of Biological Chemistry* 263.18, pp. 8929–8937.

Nielsen, S. F., Nordestgaard, B. G., and Bojesen, S. E. (2012). "Statin use and reduced cancer-related mortality." *New England Journal of Medicine* 367.19, pp. 1792–1802.

Noble, W. S. (2009). "A quick guide to organizing computational biology projects." *PLoS Computational Biology* 5.7, e1000424.

Ovaska, K., Laakso, M., Haapa-Paananen, S., Louhimo, R., Chen, P., Aittomäki, V., Valo, E., Núñez-Fontarnau, J., Rantanen, V., Karinen, S., et al. (2010). "Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme". *Genome Medicine* 2.9, p. 65.

Pérez, F. and Granger, B. E. (2007). "IPython: a System for Interactive Scientific Computing". *Computing in Science and Engineering* 9.3, pp. 21–29.

Pimplikar, S. W., Ikonen, E., and Simons, K. (1994). "Basolateral protein transport in streptolysin O-permeabilized MDCK cells". *The Journal of Cell Biology* 125.5, pp. 1025–1035.

Ramsey, N. (1994). "Literate programming simplified". *IEEE Software* 11.5, pp. 97–105.

– (2008). *Noweb—a simple, extensible tool for literate programming*. URL: `http://www.cs.tufts.edu/~nr/noweb`.

Reddy, J. V., Ganley, I. G., and Pfeffer, S. R. (2006). "Clues to neuro-degeneration in Niemann-Pick type C disease from global gene expression profiling." *PLOS ONE* 1, e19.

Reeves, V. L., Thomas, C. M., and Smart, E. J. (2012). "Lipid rafts, caveolae and GPI-linked proteins". In: *Caveolins and Caveolae*. Springer, pp. 3–13.

Rowland, A. A. and Voeltz, G. K. (2012). "Endoplasmic reticulum–mitochondria contacts: function of the junction". *Nature Reviews Molecular Cell Biology* 13.10, pp. 607–625.

Sahlberg, K. K., Hongisto, V., Edgren, H., Mäkelä, R., Hellström, K., Due, E. U., Moen Vollan, H. K., Sahlberg, N., Wolf, M., Børresen-Dale, A.-L., Perälä, M., and Kallioniemi, O. (2013). "The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells." *Mol Oncol* 7.3, pp. 392–401.

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013). "Ten simple rules for reproducible computational research." *PLoS Computational Biology* 9.10, e1003285.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). "NIH Image to ImageJ: 25 years of image analysis." *Nature methods* 9 (7), pp. 671–675.

Scotti, E., Hong, C., Yoshinaga, Y., Tu, Y., Hu, Y., Zelcer, N., Boyadjian, R., Jong, P. J. de, Young, S. G., Fong, L. G., and al., et (2011). "Targeted Disruption of the Idol Gene Alters Cellular Regulation of the Low-Density Lipoprotein Receptor by Sterols and Liver X Receptor Agonists". *Molecular and Cellular Biology* 31.9, pp. 1885–1893.

Scotti, E., Calamai, M., Goulbourne, C. N., Zhang, L., Hong, C., Lin, R. R., Choi, J., Pilch, P. F., Fong, L. G., Zou, P., Ting, A. Y., Pavone, F. S., Young, S. G., and Tontonoz, P. (2013). "IDOL stimulates clathrin-independent endocytosis and multivesicular body-mediated lysosomal degradation of the low-density lipoprotein receptor." *Molecular and Cellular Biology* 33.8, pp. 1503–1514.

Shen, H. (2014). "Interactive notebooks: Sharing the code." *Nature* 515.7525, pp. 151–152.

Sigismund, S., Argenzio, E., Tosoni, D., Cavallaro, E., Polo, S., and Di Fiore, P. P. (2008). "Clathrin-Mediated Internalization Is Essential for Sustained EGFR Signaling but Dispensable for Degradation". *Developmental Cell* 15.2, pp. 209–219.

Simons, K. and Vaz, W. L. (2004). "Model Systems, Lipid Rafts, and Cell Membranes1". *Annual Review of Biophysics and Biomolecular Structure* 33.1, pp. 269–295.

Staaf, J., Jönsson, G., Ringnér, M., Vallon-Christersson, J., Grabau, D., Arason, A., Gunnarsson, H., Agnarsson, B. A., Malmström, P.-O., Johannsson, O. T., Loman, N., Barkardottir, R. B., and Borg, A. (2010). "High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer." *Breast Cancer Research* 12.3, R25.

Strzelczyk, B., Szulc, A., Rzepko, R., Kitowska, A., Skokowski, J., Szutowicz, A., and Pawelczyk, T. (2009). "Identification of high-risk stage II colorectal tumors by combined analysis of the NDRG1 gene expression and the depth of tumor invasion." *Annals of Surgical Oncology* 16.5, pp. 1287–1294.

Sugii, S., Reid, P. C., Ohgami, N., Du, H., and Chang, T.-Y. (2003). "Distinct endosomal compartments in early trafficking of low density lipoprotein-derived cholesterol." *Journal of Biological Chemistry* 278.29, pp. 27180–27189.

Tan, X., Sun, Y., Thapa, N., Liao, Y., Hedman, A. C., and Anderson, R. A. (2015a). "LAPTM4B is a PtdIns(4,5)P2 effector that regulates EGFR signaling, lysosomal sorting, and degradation". *The EMBO Journal* 34.4, pp. 475–490.

Tan, X., Thapa, N., Sun, Y., and Anderson, R. (2015b). "A Kinase-Independent Role for EGF Receptor in Autophagy Initiation". *Cell* 160.1-2, pp. 145–160.

Trajkovic, K., Hsu, C., Chiantia, S., Rajendran, L., Wenzel, D., Wieland, F., Schwille, P., Brügger, B., and Simons, M. (2008). "Ceramide triggers budding of exosome vesicles into multivesicular endosomes." *Science* 319.5867, pp. 1244–1247.

Tsujishita, Y. and Hurley, J. H. (2000). "Structure and lipid transport mechanism of a StAR-related domain." *Nature Structural Biology* 7.5, pp. 408–414.

van der Kant, R., Zondervan, I., Janssen, L., and Neefjes, J. (2013). "Cholesterol-binding molecules MLN64 and ORP1L mark distinct late endosomes with transporters ABCA3 and NPC1." *Journal of Lipid Research* 54.8, pp. 2153–2165.

Wilhelm, L. P., Tomasetto, C., and Alpy, F. (2016). "Touché! STARD3 and STARD3NL tether the ER to endosomes." *Biochemical Society Transactions* 44.2, pp. 493–498.

Wilhelm, L. P., Wendling, C., Védie, B., Kobayashi, T., Chenard, M.-P., Tomasetto, C., Drin, G., and Alpy, F. (2017). "STARD3 mediates endoplasmic reticulum-to-endosome cholesterol transport at membrane contact sites." *The EMBO journal*.

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and Wilson, P. (2014). "Best practices for scientific computing." *PLoS Biology* 12.1, e1001745.

Yao, Q., Chen, J., Cao, H., Orth, J. D., Michael McCaffery, J., Stan, R.-V., and McNiven, M. A. (2005). "Caveolin-1 Interacts Directly with Dynamin-2". *Journal of Molecular Biology* 348.2, pp. 491–501.

Yarden, Y. and Sliwkowski, M. X. (2001). *Nature Reviews Molecular Cell Biology* 2.2, pp. 127–137.

Young, M. M., Takahashi, Y., Fox, T. E., Yun, J. K., Kester, M., and Wang, H.-G. (2016). "Sphingosine Kinase 1 Cooperates with Autophagy to Maintain Endocytic Membrane Trafficking." *Cell Rep* 17.6, pp. 1532–1545.

Yun, S. M., Yoon, K., Lee, S., Kim, E., Kong, S.-H., Choe, J., Kang, J. M., Han, T.-S., Kim, P., Choi, Y., Jho, S., Yoo, H., Bhak, J., Yang, H.-K., and Kim, S.-J. (2014). "PPP1R1B-STARD3 chimeric fusion transcript in human gastric cancer promotes tumorigenesis through activation of PI3K/AKT signaling." *Oncogene* 33.46, pp. 5341–5347.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., and Ettema, T. J. G. (2017). "Asgard archaea illuminate the origin of eukaryotic cellular complexity." *Nature* 541 (7637), pp. 353–358.

Zelcer, N., Hong, C., Boyadjian, R., and Tontonoz, P. (2009). "LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor." *Science* 325.5936, pp. 100–104.

Zhang, M., Liu, P., Dwyer, N. K., Christenson, L. K., Fujimoto, T., Martinez, F., Comly, M., Hanover, J. A., Blanchette-Mackie, E. J., and Strauss 3rd, J. F. (2002). "MLN64 mediates mobilization of lysosomal cholesterol to steroidogenic mitochondria." *Journal of Biological Chemistry* 277.36, pp. 33300–33310.

A Original publication reprints

NDRG1 functions in LDL receptor trafficking by regulating endosomal recycling and degradation

Vilja Pietiäinen[1,2,*], Boris Vassilev[1,3], Tomas Blom[1,3], Wei Wang[1,3], Jessica Nelson[4], Robert Bittman[5], Nils Bäck[1], Noam Zelcer[4] and Elina Ikonen[1,3]

[1]Institute of Biomedicine, Anatomy, University of Helsinki, Helsinki, Finland
[2]Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland
[3]Minerva Foundation Institute for Medical Research, Helsinki, Finland
[4]Department of Medical Biochemistry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Chemistry and Biochemistry, Queens College of The City University of New York, Flushing, NY, USA
*Author for correspondence (vilja.pietiainen@helsinki.fi)

Accepted 17 May 2013
Journal of Cell Science 126, 3961–3971

doi: 10.1242/jcs.128132

Summary

N-myc downstream-regulated gene 1 (NDRG1) mutations cause Charcot–Marie–Tooth disease type 4D (CMT4D). However, the cellular function of NDRG1 and how it causes CMT4D are poorly understood. We report that NDRG1 silencing in epithelial cells results in decreased uptake of low-density lipoprotein (LDL) due to reduced LDL receptor (LDLR) abundance at the plasma membrane. This is accompanied by the accumulation of LDLR in enlarged EEA1-positive endosomes that contain numerous intraluminal vesicles and sequester ceramide. Concomitantly, LDLR ubiquitylation is increased but its degradation is reduced and ESCRT (endosomal sorting complex required for transport) proteins are downregulated. Co-depletion of IDOL (inducible degrader of the LDLR), which ubiquitylates the LDLR and promotes its degradation, rescues plasma membrane LDLR levels and LDL uptake. In murine oligodendrocytes, Ndrg1 silencing not only results in reduced LDL uptake but also in downregulation of the oligodendrocyte differentiation factor Olig2. Both phenotypes are rescued by co-silencing of Idol, suggesting that ligand uptake through LDLR family members controls oligodendrocyte differentiation. These findings identify NDRG1 as a novel regulator of multivesicular body formation and endosomal LDLR trafficking. The deficiency of functional NDRG1 in CMT4D might impair lipid processing and differentiation of myelinating cells.

Key words: Endocytosis, Cholesterol, IDOL, Multivesicular body, NDRG1

Introduction

Low-density lipoprotein (LDL)-cholesterol is internalized into cells by clathrin-mediated endosytosis of the LDL receptor (LDLR) to meet cellular cholesterol needs (Goldstein and Brown, 2009; Ikonen, 2008). In endosomes, LDLR disassociates from LDL particles because of the acidic pH and returns to the plasma membrane (PM). Alternatively, recycling of the LDLR to the PM can be prevented by proprotein convertase subtilisin/kexin type 9 (PCSK9) or by the E3 ubiquitin ligase, inducible degrader of the LDLR (IDOL), which promotes ubiquitylation of the LDLR under conditions of excess cellular sterols (Zelcer et al., 2009). The ability of PCSK9 to alter LDLR levels is independent of, and additive with, the IDOL pathway and does not require LDLR ubiquitylation (Wang et al., 2012; Zhang et al., 2012). Ubiquitylation is an established signal for controlling endocytosis and sorting of membrane receptors, such as the LDLR, to multivesicular bodies (MVBs) and further to lysosomes for degradation (Mukhopadhyay and Riezman, 2007; Piper and Katzmann, 2007; Zelcer et al., 2009). From early endosomes, LDL particles are transported to MVBs/late endosomes (Goldstein and Brown, 2009) for the hydrolysis of LDL-cholesterol esters to free cholesterol (FC) by acid lipase (Sugii et al., 2003) and exit of cholesterol from the late endosomes by Niemann–Pick type C1 and 2 (NPC1 and NPC2) proteins (Carstea et al., 1997; Loftus et al., 1997).

Our interest in the possibility that N-myc downstream regulated gene 1 (NDRG1) might be involved in cellular cholesterol metabolism was initiated by Affymetrix gene expression profiling of cholesterol loaded, NPC1-deficient murine macrophages (Hölttä-Vuori et al., 2012). In this experiment, we identified NDRG1 as one of the highly induced genes in response to cholesterol loading (E. I., unpublished observations). Interestingly, a genome-wide expression analysis of NPC disease fibroblasts found that NDRG1 mRNA is also elevated in these patient-derived cells (Reddy et al., 2006). Moreover, in an RNAi screen focused on cellular lipid phenotypes, NDRG1 was identified as a gene potentially affecting cellular cholesterol levels (Bartz et al., 2009). Yet, how NDRG1 affects cholesterol homeostasis remains unknown.

NDRG1 is widely expressed, with particularly high levels in the nervous system (Lachat et al., 2002). It participates in diverse cellular functions, such as differentiation (van Belzen et al., 1997) and proliferation (Piquemal et al., 1999). NDRG1 expression is often altered in neoplasms, where it can act either as a tumor promoter or suppressor (Bandyopadhyay et al., 2003; Guan et al., 2000). In the demyelinating neuropathy Charcot-Marie–Tooth disease 4D (CMT4D), the deficiency of functional NDRG1 leads to autonomous dysfunction of Schwann cells, the glial cells of the peripheral nervous system (PNS) (Kalaydjieva

et al., 2000; Kalaydjieva et al., 1996). CMT4D can also be accompanied by abnormalities of the central nervous system (CNS) white matter (Echaniz-Laguna et al., 2007) and functional CNS disturbances (Kalaydjieva et al., 1998).

Subcellularly, NDRG1 is mostly found in the cytoplasm (Lachat et al., 2002) and in endosomes (Kachhap et al., 2007). NDRG1 can interact with proteins involved in vesicular trafficking, and was reported to regulate the recycling of E-cadherin by acting as an effector protein for Rab4a (Kachhap et al., 2007). NDRG1 also binds to prenylated Rab acceptor-1 (PRA1; also known as YIP3, RABAC1) (Hunter et al., 2005; Kim et al., 2012; King et al., 2011), which can act as a GDI displacement factor for endosomal Rab proteins, such as Rab4, Rab5, Rab7 and Rab9 (Bucci et al., 1999; Sivars et al., 2003). How these interactions contribute to the physiological function of NDRG1 or to the consequences of NDRG1 deficiency are poorly understood.

In this study, we describe a role for NDRG1 in cellular lipid homeostasis and endocytic trafficking of the LDLR. Our data further implicate NDRG1 as an important regulator of endosomal maturation that is required for efficient uptake of LDL receptor family members in both epithelial and glial cells.

Results

NDRG1 affects cellular cholesterol distribution and content

To assess whether the endosomal cholesterol content affects the abundance of the NDRG1 protein, we treated epithelial A431 cells with the hydrophobic amine U18666A. This compound induces lysosomal cholesterol accumulation, mimicking the loss of NPC1 function (Ko et al., 2001). Treatment of A431 cells with U18666A resulted in an approximately threefold increase in the endogenous NDRG1 protein content, in line with the upregulation of NDRG1 mRNA in NPC fibroblasts (Reddy et al., 2006). Moreover, loading of cells with LDL-cholesterol led to an approximately twofold increase in the NDRG1 protein levels (Fig. 1A). These findings suggest a role for NDRG1 in LDL-cholesterol trafficking.

To study whether NDRG1 affects cellular cholesterol homeostasis, we silenced NDRG1 using small interfering RNAs (siRNA) in A431 cells. NDRG1 knockdown (k/d) was assessed by immunoblotting, showing consistently a ~90% reduction (89±7% s.e.m.) in NDRG1 protein amount (supplementary material Fig. S1A,B). To examine cholesterol distribution in NDRG1 k/d cells, the cells were stained with filipin, a widely used fluorescent sterol-binding polyene antibiotic. NDRG1 silencing resulted in the redistribution of free cholesterol (FC) from the PM to perinuclear, punctate structures (Fig. 1B). Biochemical lipid analysis revealed that NDRG1 silencing was associated with a decreased cellular cholesterol content. Specifically, while FC remained essentially unchanged upon NDRG1 k/d, an approximately fourfold reduction was observed in cholesterol esters (CE; Fig. 1C). No major changes in cholesterol biosynthesis or efflux to apolipoprotein A-I (apoA-I) were observed in NDRG1 k/d cells (the ratio of values in NDRG1-silenced cells to control cells was 1.26±0.11 and 1.22±0.06, respectively). These data indicate that NDRG1 depletion leads to an abnormal endosomal distribution of FC. Furthermore, as cholesterol biosynthesis and efflux were not substantially affected by NDRG1 k/d, the reduction in the CE content could result from decreased cellular cholesterol uptake.

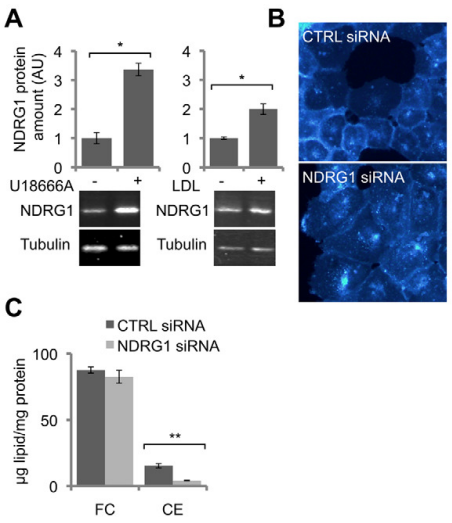


Fig. 1. NDRG1 is regulated by endosomal cholesterol loading and its depletion affects cholesterol distribution and amount. (**A**) U18666A or LDL loading increases cellular NDRG1 protein content. A431 cells were treated with U18666A (1 µg/ml) in 10% FBS for 48 hours or with LDL (50 µg/ml) in serum-free medium for 18 hours, followed by western blotting of NDRG1 and β-tubulin as a loading control. n=4–6 samples; values are means ± s.e.m., *P<0.05 (Student's t-test). (**B**) Cholesterol distribution is affected by NDRG1 silencing in A431 cells. Cells were transfected with control (CTRL) or NDRG1 siRNAs for 72 hours, fixed, and stained with filipin. (**C**) Cholesterol ester (CE) content is decreased upon NDRG1 k/d. After siRNA treatment of A431 cells, the lipids were extracted and analyzed by TLC. Free cholesterol (FC) and CEs were quantified by densitometry. Results of a representative (means ± s.d.) of three independent experiments with comparable results are shown. **P<0.001 (Student's t-test).

Decreased LDLR plasma membrane localization and degradation leads to diminished LDL uptake in NDRG1-depleted cells

To test whether NDRG1 depletion affects LDL uptake, we monitored the binding and internalization of 1,1′-dioctadecyl-3,3,3′,3′-tetramethyl-indocarbocyanine-perchlorate-labeled LDL (DiI–LDL) in NDRG1 siRNA-transfected A431 cells. LDL uptake was visualized by fluorescence microscopy and quantified by measuring the fluorescence intensity in cell lysates. These assays revealed that both the binding and internalization of DiI–LDL were severely (up to ~50%) diminished upon NDRG1 k/d (Fig. 2A; see also supplementary material Fig. S1C). This phenotype could, in principle, result from decreased LDLR levels. Unexpectedly, the total LDLR protein content was found to be increased upon NDRG1 silencing (Fig. 2B; see also supplementary material Fig. S1B). Biotinylation of the PM pool of LDLR, revealed, however, a clear decrease in the fraction of LDLR in the PM of NDRG1 k/d cells (Fig. 2B). To follow the intracellular degradation of PM LDLR, the biotinylated receptors were internalized for 0–5 hours, pulled down by streptavidin beads, and detected with an anti-LDLR antibody by immunoblotting. In control-siRNA-treated cells, ~80% of the biotinylated LDLR was degraded in 5 hours. However, in NDRG1 k/d cells, degradation of the LDLR

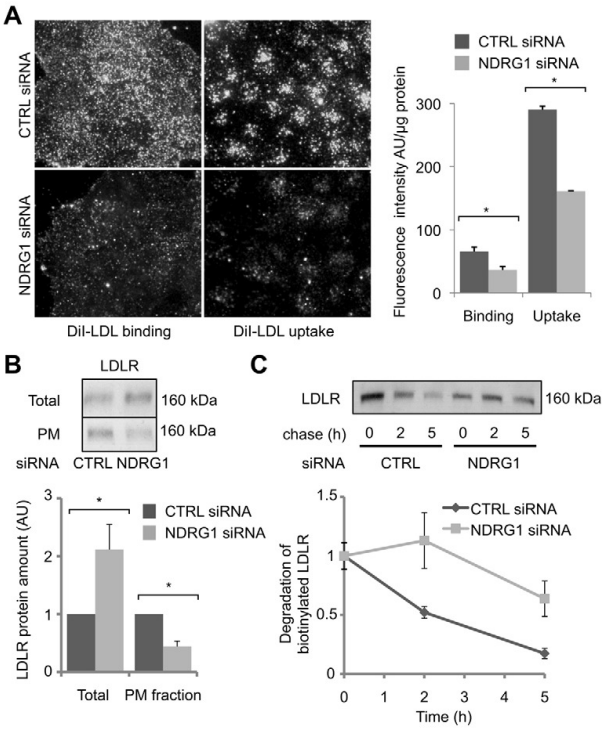


Fig. 2. NDRG1 depletion results in diminished LDL uptake, and decreased PM localization and degradation of LDLR. (**A**) NDRG1 depletion decreases the binding and uptake of DiI–LDL. siRNA-transfected A431 cells were incubated with 5 µg/ml DiI–LDL for 30 minutes at 4℃ (binding) or 20 minutes at 37℃ (uptake), and imaged with a wide-field microscope immediately post-fixation. Cellular fluorescence intensity was quantified from cell lysates, and normalized to the protein amounts (right panel). Values are means ± s.e.m, *$P<0.05$ (Student's *t*-test). (**B**) Total LDLR is increased but the PM fraction of LDLR is decreased in NDRG1 k/d cells. siRNA-transfected A431 cells were subjected to cell-surface biotinylation and streptavidin pull-down. LDLR amounts were analyzed by western blotting and quantified by densitometry; Total, input protein; PM, pull-down of biotinylated PM protein. Values are means ± s.e.m. of three to five experiments, *$P<0.05$ (Student's *t*-test). (**C**) LDLR degradation is slowed down in NDRG1-silenced cells. After surface biotinylation, cells were incubated for the indicated times at 37℃ to allow internalization and degradation of biotinylated LDLR, before analysis as in B. Please note that PM LDLR levels at 0 hours were set as 1; in CTRL-siRNA-treated samples there was an approximate twofold more LDLR at the PM than in NDRG1-siRNA-treated samples. Values are means ± s.e.m. of three samples/condition from two independent experiments.

was severely attenuated with less than 40% of the receptor lost in 5 hours (Fig. 2C). Together, these results demonstrate that NDRG1 is required for the proper PM localization of the LDLR. The reduction of LDLR at the PM results in diminished LDL-cholesterol uptake and is accompanied by retarded degradation of the LDLR.

Silencing of NDRG1 or its interaction partner PRA1 leads to endosomal LDLR sequestration

To define the intracellular localization of excess LDLRs in NDRG1-silenced cells, LDLR was co-immunostained with markers for early endosomes (endosomal antigen 1; EEA1) and late endosomes/lysosomes (Lysosome-associated membrane glycoprotein 1; LAMP1) and imaged by confocal microscopy. In control cells, LDLR typically localized to small, punctate EEA1-positive endosomes in the perinuclear region and to some extent to LAMP1-positive late endosomes/lysosomes (Fig. 3A). In NDRG1-silenced cells, enlarged endosomes and lysosomes were more irregularly scattered in the cytoplasm. Strikingly, the LDLR was largely sequestered in the abnormal EEA1-positive endosomal vesicles (Fig. 3A). These structures also accumulated FC, as detected by filipin staining (supplementary material Fig. S2).

To investigate how NDRG1 deficiency results in the strikingly altered endosome morphology, we considered the involvement of the previously described NDRG1 interaction partners Rab4a and PRA1 (Hunter et al., 2005; Kachhap et al., 2007; Kim et al., 2012; King et al., 2011). Therefore, we tested the consequences of Rab4a or PRA1 silencing on LDLR distribution by immunostaining of A431 cells. We observed that the LDLR phenotype of Rab4a depletion was not identical to that of NDRG1 depletion, with the latter leading to larger intracellular LDLR-positive accumulations (Fig. 3B). However, depletion of PRA1 led to the appearance of enlarged early endosomes and late endosomes (Fig. 3C) as well as to the accumulation of the LDLR in these enlarged EEA1-positive endosomes, closely resembling the LDLR phenotype of NDRG1 k/d cells. We then analyzed whether NDRG1 silencing affects the levels of endogenous PRA1 or endosomal Rab proteins in A431 cells. Remarkably, we observed a significant decrease in the amount of endogenous PRA1 protein, both by immunostaining and immunoblotting of NDRG1 k/d cells (Fig. 4A; see also supplementary material Fig. S3). In contrast, the levels of Rab4, 5 and 7 were not substantially altered (supplementary material Fig. S4).

To address whether the reduced levels of PRA1 contribute to the LDLR trafficking defect observed upon NDRG1 depletion, we overexpressed PRA1 in NDRG1-depleted cells. After fixation, the cells were immunostained for PRA1 and LDLR and assessed for the degree of LDLR endosomal sequestration. pEGFP was expressed as a control to evaluate the possible effect of transfection on LDLR localization. In control cells overexpressing GFP, LDLR showed a punctate staining throughout the cytoplasm, whereas in NDRG1 k/d cells overexpressing GFP, LDLR was abnormally sequestered as expected (Fig. 4B). Importantly, overexpression of PRA1 partially rescued the LDLR phenotype in NDRG1-siRNA-treated

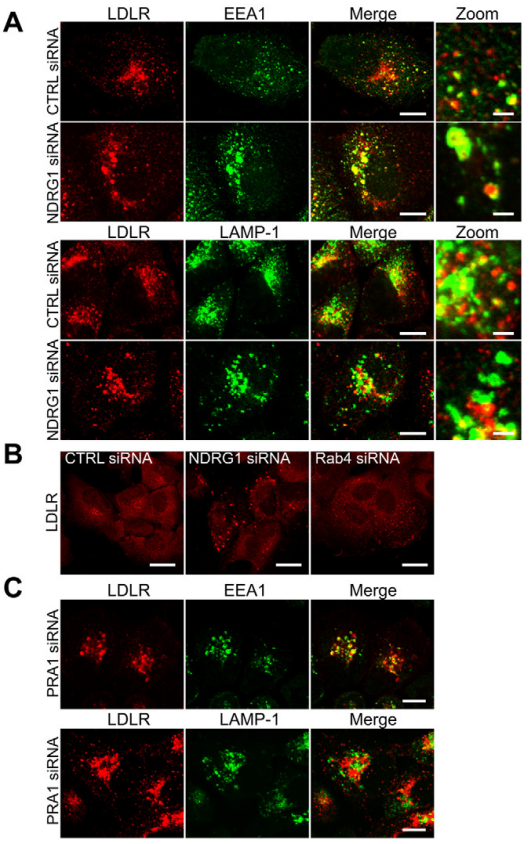


Fig. 3. LDLR accumulates in modified enlarged endosomes upon knockdown of NDRG1 or PRA1. (**A**) LDLR localizes in enlarged endosomes in NDRG1-depleted cells. Confocal images of siRNA-transfected A431 cells, immunostained for LDLR (red), EEA1 (green) and LAMP-1 (green). Scale bars:10 μm (Merge), 5 μm (Zoom). (**B**) Silencing of Rab4a interferes with LDLR distribution in A431 cells but the phenotype differs from that of NDRG1 silencing. Cells were permeabilized with filipin and immunostained with LDLR antibody. Scale bars: 20 μm. (**C**) PRA1 silencing leads to accumulation of LDLR in enlarged EEA1-positive endosomes. Confocal images of PRA1 siRNA-transfected A431 cells, immunostained for LDLR (red) and EEA1 (green) or LAMP1 (green). Scale bars: 10 μm.

cells (Fig. 4B). These results demonstrate that NDRG1 deficiency leads to an abnormal endosomal sequestration of the LDLR. They also suggest that the reduced levels of PRA1 accompanying NDRG1 deficiency might contribute to the aberrant endosomal morphology and LDLR sequestration.

NDRG1 affects the morphology and function of multivesicular bodies

The effect of NDRG1 silencing on the ultrastructure of endosomal organelles was studied by electron microscopy (EM). To this end, siRNA-transfected A431 cells were allowed to internalize BSA-gold for 40 min to label the endosomes and subsequently processed for EM. Both in the control- and NDRG1-silenced cells, BSA-gold accumulated in vesicles with MVB-like morphology (Fig. 5A). For quantification, we measured the diameter of 60 MVBs from unlabeled siRNA-treated EM samples and counted the number of intraluminal vesicles in each. Compared with control cells, MVBs in NDRG1-silenced cells were enlarged and contained approximately threefold more intraluminal vesicles (Fig. 5A).

Because the formation of intraluminal vesicles destined for lysosomal degradation requires endosomal sorting complex required for transport (ESCRT) proteins (Raiborg and Stenmark, 2009), we assessed the levels of the ESCRT components in NDRG1-silenced cells by western blotting. The immunoblots revealed a significant reduction in the amounts of the tested ESCRT proteins, i.e. Hrs (ESCRT-0), TSG101 (ESCRT-I), Vps22 (ESCRT-II), Vps24 (ESCRT-III) and Vps32 (ESCRT-III) (Fig. 5B). In addition to ESCRTs, lipids might play an important role in the membrane invagination step of MVB formation. MVBs are abundant in sphingolipids, such as sphingomyelin (SM) and ceramide, and the latter can induce the formation of internal vesicles in liposomes and trigger budding of exosome vesicles into multivesicular endosomes in cells (Trajkovic et al., 2008). Interestingly, we found a significant increase in the total ceramide levels of NDRG1 k/d cells (Fig. 5C).

To test whether the increase in ceramide is due to its accumulation in the enlarged MVBs, we incorporated a fluorescent derivative of ceramide (BODIPY-labeled ceramide) into LDL particles, and applied them to NDRG1-silenced cells. The cells had been prelabeled overnight with the fluorescent fluid-phase tracer dextran to visualize lysosomes. (Please note that dextran distribution is altered upon NDRG1 depletion due to the redistribution of late endocytic organelles; see supplementary material Fig. S5.) As shown in Fig. 6A, after 1 hour of LDL labeling BODIPY-labeled ceramide was already partially colocalized with dextran in control cells. The extent of colocalization increased until a 2-hour chase, after which it decreased (Fig. 6A). However, in NDRG1 k/d cells, BODIPY-labeled ceramide localized to a lesser extent with dextran after 1 hour of labeling and remained for a prolonged time in enlarged dextran negative vesicular structures (Fig. 6A). This suggests that the transport of ceramide through the endosomal pathway is slower in NDRG1-depleted cells. To address the source of increased endosomal ceramide, we internalized LDL particles labeled with [^{3}H]SM into cells. This revealed increased levels of [^{3}H]SM-derived [^{3}H]ceramide in NDRG1-silenced cells, providing evidence that at least part of the ceramide derives from endosomal SM hydrolysis (Fig. 6B). Together, these findings imply that NDRG1 function is required for proper endosomal maturation. In NDRG1 deficiency, MVBs accumulate increased numbers of intraluminal vesicles despite a marked reduction of ESCRT components. Furthermore, the endosomal retention of both fluorescent ceramide and increased LDL/[^{3}H]SM-derived [^{3}H]ceramide suggests that the intraluminal vesicles in NDRG1-depleted cells are ceramide enriched.

Prevention of IDOL-induced LDLR ubiquitylation rescues LDLR distribution and cholesterol levels in NDRG1-silenced cells

Ubiquitylation of membrane proteins can negatively regulate their cell surface expression by targeting them to the MVBs and for degradation (Staub and Rotin, 2006). Interestingly, silencing

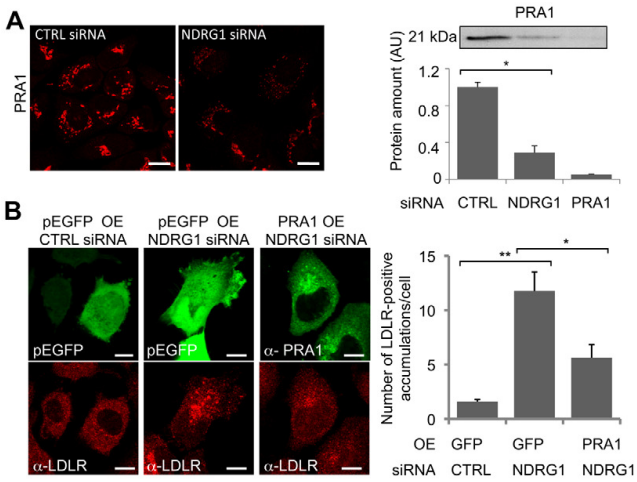


Fig. 4. PRA1 is reduced in NDRG1-silenced cells and its overexpression rescues LDLR distribution in these cells. (**A**) NDRG1 silencing is accompanied by reduced amounts of PRA1. Immunostaining (confocal images) and western blotting of endogenous PRA1 in siRNA-transfected A431 cells. Values are means ± s.e.m. of two to three experiments with parallel samples; *$P<0.05$ (Student's *t*-test). Scale bars: 20 µm. (**B**) PRA1 overexpression (OE) rescues the LDLR accumulation phenotype of NDRG1-silenced cells. PRA1 was overexpressed in NDRG1-silenced A431 cells. Cells overexpressing PRA1 were detected with anti-PRA1 antibody (green), and endogenous LDLR was detected with anti-LDLR antibody (red). pEGFP was used as a transfection control. The number of intracellular LDLR-positive accumulations was quantified from confocal images, $n=70$ cells/sample; values are means ± s.e.m., *$P<0.05$; **$P<0.001$ (Student's *t*-test). Scale bars: 10 µm.

of NDRG1 enhanced the colocalization of the LDLR with ubiquitin in enlarged endosomes (Fig. 7A). This suggests that the abnormal endosomes in NDRG1-silenced cells can accumulate ubiquitin and possibly also ubiquitylated LDLR despite reduced ESCRT levels. To directly analyze whether NDRG1 silencing affects the ubiquitylation status of the LDLR, we used HEK293T cells. The cells were co-transfected with control or NDRG1 siRNAs and an expression plasmid encoding HA-tagged LDLR. The cell lysates were immunoprecipitated with anti-HA beads, followed by immunoblotting with an anti-ubiquitin antibody to detect the ubiquitylated LDLR. As shown in Fig. 7B, depletion of NDRG1 results in increased amounts of ubiquitylated LDLR.

Recently, the E3 ubiquitin ligase IDOL was identified as a novel post-translational regulator of LDLR ubiquitylation (Hong et al., 2010; Sorrentino and Zelcer, 2012; Zelcer et al., 2009). We examined whether prevention of LDLR ubiquitylation by IDOL

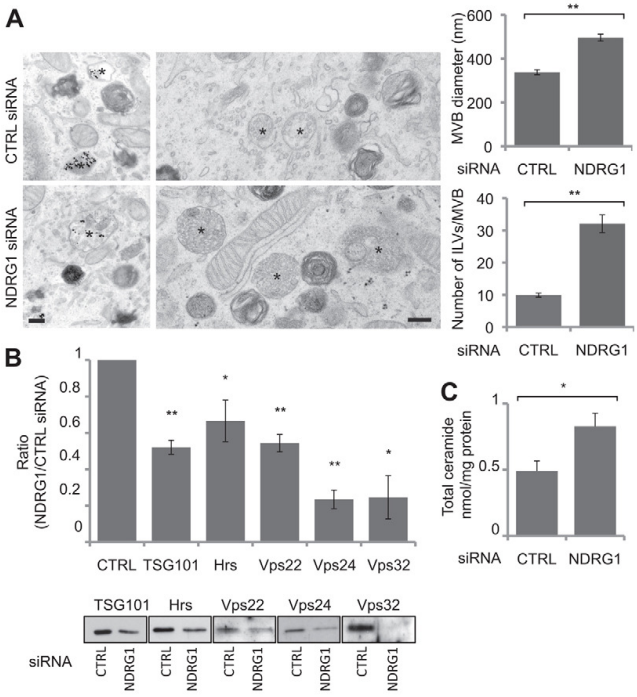


Fig. 5. NDRG1 affects MVB morphology and ESCRT levels. (**A**) NDRG1 silencing results in enlarged MVBs with increased number of intraluminal vesicles. Left: EM images of siRNA-transfected A431 cells, labeled with BSA-gold for 40 minutes. MVBs with ILVs are marked by an asterisk. MVB diameter and the number of intraluminal vesicles (ILVs) were quantified from 60 MVBs of unlabeled EM samples. Values are means ± s.e.m., **$P<0.001$ (Student's *t*-test). Scale bars: 200 nm. (**B**) NDRG1 silencing results in the downregulation of ESCRTs. ESCRT protein amounts relative to CTRL were analyzed by western blotting of siRNA-transfected A431 cells. The representative immunoblots are shown for each ESCRT protein. Values are means ± s.e.m. from two to five independent experiments, *$P<0.05$, **$P<0.001$ (Student's *t*-test). (**C**) Total ceramide content is increased upon NDRG1 depletion in A431 cells. Total cellular ceramide amounts were determined in siRNA-treated A431 cells by HPTLC and densitometry. Values are means ± s.e.m. of five samples/condition from two independent experiments, *$P<0.05$ (Student's *t*-test).

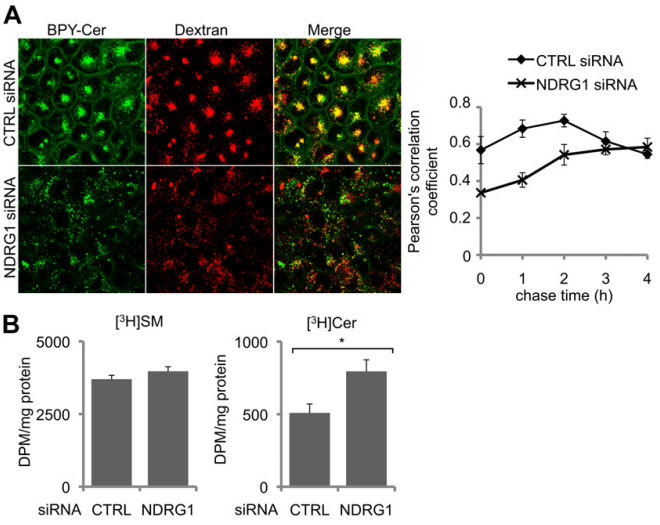


Fig. 6. Ceramide is enriched in enlarged MVBs in NDRG1 knockdown cells. (**A**) Altered BODIPY-labeled ceramide (BPY-Cer) transport in NDRG1-silenced cells. Cells pre-labeled with dextran were pulse-labeled with BPY-Cer/LDL for 1 hour and chased for 0–4 hours (images from 1 hour chase). Colocalization of BPY-Cer (green) and dextran (red) was analyzed using Pearson's correlation coefficient at 0–4 hours of chase. $n=3$ confocal sections and >30 cells per time point; values are means ± s.e.m. (**B**) [^{3}H]Ceramide accumulation derives from endosomal SM hydrolysis in NDRG1-silenced cells. [^{3}H]SM incorporated into LDL particles was internalized into A431 cells for 2 hours and the amounts of [^{3}H]SM and derived [^{3}H]ceramide were analyzed by TLC. Values are means ± s.e.m. of six samples/condition from two independent experiments, $*P<0.05$ (Student's t-test).

depletion could rescue the LDLR levels at the PM and reverse the endosomal LDLR accumulation in NDRG1-depleted cells. As expected, silencing of IDOL led to an increase in both total and PM LDLR abundance (Fig. 7C). Remarkably, depletion of IDOL in NDRG1 k/d cells was able to reverse the decrease in LDLR PM levels (Fig. 7C). Furthermore, under these conditions, depletion of IDOL alleviated the endosomal sequestration of LDLR (Fig. 7D). We also assessed the consequence of IDOL silencing on cellular cholesterol stores. Silencing of IDOL resulted in increased CES, whereas silencing of NDRG1 led to the opposite outcome (Fig. 7E). Strikingly, when IDOL and NDRG1 were silenced simultaneously, the level of CEs was rescued (Fig. 7E).

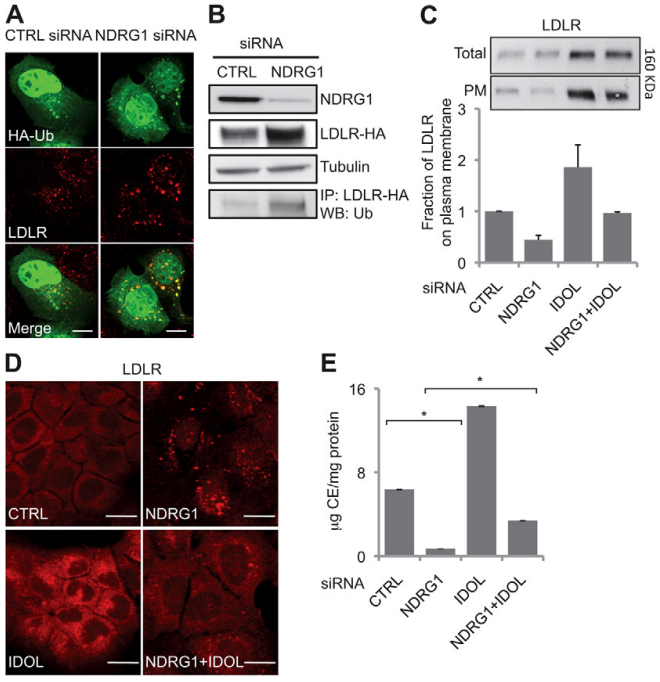


Fig. 7. LDLR ubiquitylation is increased in NDRG1-silenced cells, and its prevention by IDOL depletion rescues LDLR distribution and CE content. (**A**) LDLR colocalizes with HA–Ub in NDRG1-depleted cells. A431 cells were transfected with siRNAs and a plasmid encoding HA–Ub. HA–Ub was immunostained with anti-HA antibodies (green) and LDLR with anti-LDLR antibodies (red). Confocal microscopy images are shown. Scale bars: 10 µm. (**B**) LDLR is heavily ubiquitylated upon NDRG1 k/d. HEK293T cells were co-transfected with LDLR-HA and siRNAs and immunoblotted as indicated. LDLR–HA was immunoprecipitated from total cell lysates using HA-beads and the ubiquitylated LDLR–HA was detected with an anti-Ub antibody; α-tubulin was used as a loading control. A representative immunoblot of three independent experiments with comparable results is shown. (**C**) IDOL k/d rescues the PM localization of LDLR in NDRG1-silenced cells. siRNA-transfected A431 cells were analyzed by surface biotinylation as in Fig. 2B. Values are means ± s.e.m. of two independent experiments; a representative immunoblot is shown. (**D**) Depletion of IDOL in NDRG1-silenced cells complements the endosomal LDLR accumulation. siRNA-transfected A431 cells were permeabilized with filipin, immunostained with anti-LDLR antibody and imaged by confocal microscopy, using a 40× objective. Scale bars: 20 µm. (**E**) Depletion of IDOL in NDRG1-silenced cells rescues cellular CE content. To determine the CEs in siRNA-treated cells, lipids were extracted, separated by TLC, and quantified by densitometry. Values are means ± s.e.m., $*P<0.05$ (Student's t-test).

Taken together, these results demonstrate that LDLR is ubiquitylated in NDRG1-depleted cells and accumulates in abnormal enlarged endosomes/MVBs. If LDLR ubiquitylation in these cells is prevented, the receptor is not similarly sequestered in endosomes but can apparently recycle to the PM to rescue LDLR PM levels and LDL uptake. Accordingly, the cholesterol content of NDRG1-depleted cells is rescued.

NDRG1-depletion reduces glial cell LDL uptake and decreases the amount of Olig2 in an Idol-dependent manner

NDRG1 is highly expressed in myelinating glial cells (Okuda et al., 2008) and *NDRG1* mutations cause CMT4D. The major CMT4D-causing founder mutation of NDRG1 contains a premature termination signal at residue 148 (R148X) (Kalaydjieva et al., 2000). The *Ndrg1*$^{-/-}$ mice display a progressive demyelinating disorder, providing evidence that NDRG1 deficiency accounts for the disease (Okuda et al., 2004). Accordingly, we observed that although the overexpression of WT NDRG1 rescued the LDLR accumulation in NDRG1 k/d cells (supplementary material Fig. S6), a construct of NDRG1 truncated at R148 was unstable and was unable to rescue the phenotype (V. P. and E. I., unpublished observations). These data suggest that the effects of NDRG1 silencing might mimic the effects of functional NDRG1 deficiency in CMT4D.

We therefore studied the role of Ndrg1 in regulating LDL uptake in murine Oli-neu cells, a widely used oligodendroglial cell line generated by immortalization of mitotic oligodendrocyte precursor cells by retroviral *t-neu* oncogene expression (Jung et al., 1995). Silencing of Ndrg1 in Oli-neu cells (see supplementary material Fig. S1D) reduced DiI–LDL binding and uptake (Fig. 8A), analogously to our observations in A431 cells. In line with the effect of IDOL in other cell types, silencing of Idol in Oli-neu cells resulted in an increase in both DiI–LDL binding and uptake (Fig. 8A). Furthermore, depletion of Idol in Ndrg1-silenced cells re-established DiI–LDL binding and uptake to the level of control-siRNA-treated cells (Fig. 8A). These results suggest that Ndrg1 and Idol are involved in LDL uptake in oligodendrocytes and that these processes are interconnected.

Light microscopy revealed that control-siRNA-treated Oli-neu cells exhibited outgrowths characteristic of oligodendrocytes, whereas the Ndrg1-depleted cells had a distinct, spindle-shaped morphology. To quantify the number of outgrowths, control and Ndrg1-siRNA-treated cells were transfected with an expression plasmid encoding a constitutively farnesylated GFP construct (GFP-F). This construct localizes to the PM and enables visualization of neurite outgrowths (Lalli and Hall, 2005). The Ndrg1-depleted cells had fewer outgrowths per cell than control cells (Fig. 8B). A possible explanation is that the silencing of Ndrg1 affects the differentiation of Oli-neu cells. To test this idea, we immunoblotted and immunostained siRNA-transfected Oli-neu cells with an antibody against oligodendrocyte lineage transcription factor 2 (Olig2). Olig2 is necessary for the genesis of oligodendrocytes and is an important regulator of oligodendrocyte differentiation and myelination (Ligon et al., 2006). In Ndrg1-depleted cells Olig2 immunoreactivity was significantly reduced (~50%, based on immunoblotting; Fig. 8C) and the remaining Olig2 protein was largely cytoplasmic, in contrast to control cells that showed a strong nuclear Olig2 localization (V.P. and E.I., unpublished observations).

Finally, we examined the potential connection of oligodendrocyte differentiation and LDLR regulation by analyzing Olig2 levels after Idol k/d. Depletion of Idol slightly increased Olig2 protein levels. Importantly, Idol co-depletion in Ndrg1-silenced cells was able to rescue the reduced Olig2 levels (Fig. 8C). Cumulatively, these findings suggest that NDRG1 is

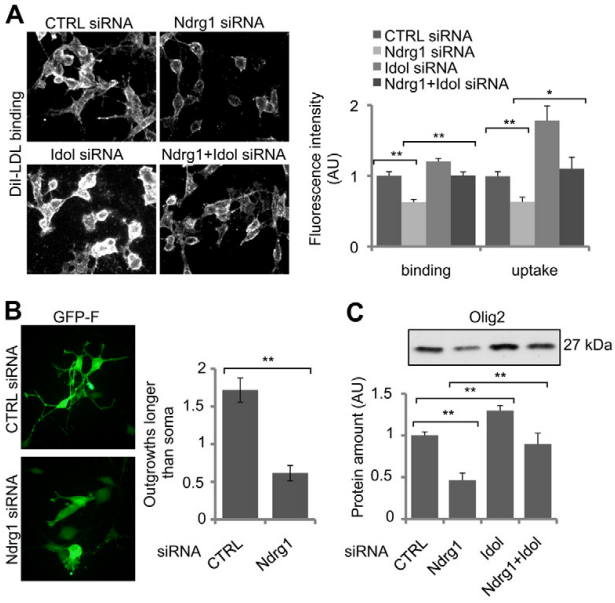


Fig. 8. Ndrg1 knockdown in murine oligodendrocytes results in reduced LDL uptake and decreases the amount of Olig2. (**A**) DiI–LDL binding and uptake are reduced in Ndrg1-silenced Oli-neu cells. Cells were incubated with 2 μg/ml DiI–LDL for 20 minutes at 4℃ (binding) or for 30 minutes at 37℃ (uptake). Cells were imaged and the fluorescence intensity per cell was quantified. Values are means ± s.e.m. of two to four independent experiments (at least 900 cells analyzed/condition); $^{**}P<0.001$, $^{*}P<0.05$ (Student's *t*-test). (**B**) Number of Oli-neu cell outgrowths is diminished upon Ndrg1 k/d. A plasmid encoding GFP-F was transfected to siRNA-treated Oli-neu cells to visualize and quantify outgrowths. Values are mean ± s.e.m., 50 cells/condition, $^{**}P<0.001$ (Student's *t*-test). (**C**) Ndrg1 silencing leads to a reduction in the amount of Olig2 but co-depletion of Idol complements the phenotype. Oli-neu cells were collected after 72 hours of siRNA transfection and Olig2 protein amounts were analyzed by western blotting and quantified by densitometry. Values are means ± s.e.m. of four experiments, $^{**}P<0.001$ (Student's *t*-test).

important for oligodendrocyte cholesterol balance and differentiation and that these actions might be mediated through IDOL-sensitive LDLR family members (Hong et al., 2010).

Discussion

NDRG1 as a regulator of LDLR trafficking

When cells are in need of cholesterol, the endocytosed LDLR normally recycles rapidly to the PM (Goldstein and Brown, 2009). We found that in NDRG1-depleted cells, the receptor was reduced at the cell surface and accumulated in modified early endosomes and MVBs. In parallel, the degradation of LDLR was slowed down, contributing to an increase in the total cellular LDLR. The decreased LDLR degradation is in agreement with the fact that a functional MVB pathway is needed for lysosome biogenesis and is essential for cargo degradation (Mukhopadhyay and Riezman, 2007).

LDLR ubiquitylation by IDOL is known to act as a sorting signal that directs the receptor for lysosomal degradation (Hong et al., 2010; Zelcer et al., 2009). Our findings indicate that the LDLR in NDRG1-depleted cells is heavily ubiquitylated, suggesting that it is targeted for lysosomal degradation. We propose that the LDLR must be ubiquitylated by IDOL in order to be sequestered into the abnormal endosomal compartment observed in NDRG1-deficient cells. In line with this, LDLR is able to recycle to the PM in the simultaneous absence of IDOL and NDRG1. This recycling LDLR is likely to utilize distinct endosomal subpopulations or subdomains from those involved in MVB formation and eventual cargo degradation. Indeed, endosomes are known to exist in cargo-specific populations (Lakadamyali et al., 2006) and cargo-specific signals can modify the endosomal pathway (Miaczynska et al., 2004; van der Goot and Gruenberg, 2006).

NDRG1 as a regulator of endosomal morphology and function

The enlarged MVBs in NDRG1 k/d cells contained more intraluminal vesicles, suggesting that the mechanism(s) of intraluminal vesicle formation and/or turnover were altered upon NDRG1 depletion. These atypical MVBs sequestered recycling membrane proteins, such as the LDLR. Of note, they also accumulate the transferrin receptor, E-cadherin (Kachhap et al., 2007), and β1-integrin (V. P. and E. I., unpublished observations). In addition, they accumulated lipids (free cholesterol and ceramide) and ubiquitin, and were positive for EEA1. The mechanisms by which the enlarged MVBs and increased number of intraluminal vesicles are generated upon NDRG1 k/d might involve several mechanisms. Based on earlier studies and our observations these mechanisms could include: (1) altered activity of endosomal Rab GTPases (Huotari and Helenius, 2011); (2) downregulation of ESCRT complex components (Stuffers et al., 2009); (3) increased cholesterol/ceramide content in MVBs (Kobuna et al., 2010; Trajkovic et al., 2008); (4) increased ubiquitylation of cargo (MacDonald et al., 2012); and (5) decreased degradation of cargo (Huotari and Helenius, 2011).

The biogenesis of the atypical MVBs in NDRG1 k/d cells was accompanied by diminished amounts of PRA1 and ESCRT complex components. Of PRA1, about two-thirds was depleted upon NDRG1 silencing. PRA1 has been shown to interact with NDRG1 and regulate several endosomal Rab GTPases, including Rabs 4, 5, 7 and 9 (Bucci et al., 1999; Sivars et al., 2003), and to affect the endosomal cholesterol content (Liu et al., 2011). Moreover, we found that PRA1 overexpression alleviates the LDLR redistribution in NDRG1-silenced cells. Therefore, the shortage of PRA1 in NDRG1-silenced cells might alter the delivery of Rab GTPases to endosomal membranes (Sivars et al., 2003), thereby contributing to the disturbed endosomal morphology and membrane transport.

Interestingly, our data indicate that MVB biogenesis and endosomal concentration of ubiquitylated cargo can proceed in NDRG1-depleted cells despite downregulation of ESCRT proteins. Although ESCRT-independent mechanisms for MVB formation have been proposed (Stuffers et al., 2009; Trajkovic et al., 2008), the physiological regulators of this process remain largely unknown. We suggest that in NDRG1-silenced cells, the endosomal accumulation of ceramide (presumably deriving from SM hydrolysis) triggers the formation of intraluminal vesicles. As such, NDRG1 appears as the first protein to negatively regulate the formation of ceramide-rich intraluminal vesicles. In future studies, it will be important to elucidate in more detail how NDRG1 could control this process.

NDRG1 and CMT4D

To extend our findings to a pathophysiologically relevant cell model, we studied glial cells. We show that in rodent oligodendrocytes, which normally exhibit strong Ndrg1 expression in the cytoplasm and processes (Berger et al., 2004; Okuda et al., 2008), Ndrg1 deficiency results in reduced LDL uptake. Importantly, this implies that the NDRG1 effect on LDL uptake is not specific for a certain cell type or species. High cholesterol availability is known to be required for myelin membrane growth (Saher et al., 2005). In addition, studies in other CMT neuropathies have emphasized the importance of endocytic pathways in the formation of myelin (Lee et al., 2012; Roberts et al., 2010; Sidiropoulos et al., 2012; Stendel et al., 2010; Verhoeven et al., 2003), reinforcing a link between NDRG1 function in endocytic trafficking and myelin formation/maintenance.

NDRG1 might also affect oligodendrocyte differentiation, as suggested by the lack of outgrowths and downregulation of Olig2 in Ndrg1-silenced oligodendrocytes. This agrees with the notion that the NDRG1 transcript is strongly upregulated during oligodendrocyte differentiation (Cahoy et al., 2008). We describe for the first time that Idol is involved in the regulation of lipid uptake in oligodendroglial cells. This finding is consistent with a recent report showing that liver X receptors (LXRs) that are transcriptional regulators of Idol expression control lipid metabolism in this cell type (Nelissen et al., 2012).

Intriguingly, IDOL not only promotes the ubiquitylation of the LDLR (Zelcer et al., 2009) but also two related receptors, the ApoE receptor 2 (ApoER2) and very-low-density lipoprotein receptor (VLDLR) (Hong et al., 2010), known to be expressed in oligodendrocytes (Zhao et al., 2007). Indeed, a potential explanation for the rescue of Olig2 levels by Idol depletion in Ndrg1 k/d cells is that uptake of LDL, or other ligands of IDOL-regulated receptors, is required to maintain Olig2 expression and oligodendrocyte differentiation. Notably, both ApoER2 and VLDLR have been shown to mediate Idol-sensitive Reelin signaling (Hong et al., 2010), a pathway crucial for the proper development of the nervous system. Moreover, Olig2-expressing oligodendroglial precursor cells are absent in mice lacking

megalin, another LDLR family member involved in the endocytosis of multiple ligands (Wicher and Aldskogius, 2008).

In summary, this study identifies NDRG1 as a novel regulator of MVB integrity and LDLR endosomal recycling. In the absence of this protein, ubiquitylated LDLR accumulates in modified early endosomes/MVBs that harbor increased amounts of ceramide-rich intraluminal vesicles. Thus, NDRG1 emerges as a critical protein that stimulates receptor recycling and limits the number of intraluminal vesicles in MVBs. Exactly how NDRG1 co-operates with known regulators of this process, such as ESCRT components and Rab GTPases, and how broadly NDRG1 regulates recycling membrane proteins, remain to be addressed. Importantly, the regulatory function of NDRG1 in endosomal dynamics appears to hold for several cell systems; therefore, it is likely to be relevant for understanding the contribution of NDRG1 both in neuropathy and in tumorigenesis.

Materials and Methods

Antibodies and other reagents

The following antibodies were used: NDRG1 (rabbit pAb; T. Commes, Universite Montpellier, France), NDRG1 (rabbit pAb; Atlas Antibodies AB, Stockholm, Sweden), LDLR (mAb C7, mouse hybridoma supernatant; Beisiegel et al., 1981; immunostainings), LDLR (rabbit pAb, 20R-LR002, Fitzgerald Industries International, Acton, Massachusetts, USA; immunostainings and western blotting), LDLR (rabbit mAb, clone EP1553Y, Epitomics, Burlingame, CA, USA; western blotting), LAMP1 (mAb H4A3, Developmental Studies Hybridoma Bank, Iowa, Maryland, USA), EEA1 (mAb, BD Biosciences, San Jose, CA, USA), PRA1 (mAb 2A4, Novus Biologicals, Littleton, Colorado, USA), TSG101 (mAb 4A10, GeneTex Inc., Irvine, CA, USA), ubiquitin (mAb FK2, BML-PW8810, Enzo Life Sciences, Farmingdale, New York, USA), hemagglutinin (HA) (mAb, Sigma-Aldrich, St. Louis, MO, USA), β-tubulin (mAb, T4026, Sigma-Aldrich), α-tubulin (mAb, T9026/clone DM1A, Sigma) and Hrs, Vps22, Vps24 and Vps32 rabbit pAbs (H. Stenmark). U18666A [3β-(2-diethylaminoethoxy)androst-5-en-17-one], HPLC-grade solvents, lovastatin, and silica gel 60 TLC plates were from Merck (Darmstadt, Germany), [1,2-^{3}H]cholesterol from Perkin Elmer (Waltham, MA), [^{3}H]acetic acid from GE Healthcare (Chalfont St. Giles, UK), dextran–Alexa-Fluor-568 (10000MW) and DiI–LDL from Invitrogen-Life Technologies (Carlsbad, California, USA), Effectene and Hiperfect from Qiagen (Germantown, Maryland, USA), protease inhibitors (chymostatin, leupeptin, antipain and pepstatin A), Triton X-100, filipin and mevalonic acid lactone (mevalonate) from Sigma-Aldrich and diethyl ether and petroleum ether from J. T. Baker (Avantor Performance Materials, Center Valley, Pennsylvania, USA). For electron microscopy, glutaraldehyde and osmium tetroxide were purchased from Electron Microscopical Sciences (Hatfield, Pennsylvania, USA) and sodium cacodylate and potassium ferrocyanide were from Sigma-Aldrich. D-erythro-*N*-palmitoyl-BODIPY-sphingosine (Ceramide-BDY) was prepared by *N*-acylation of BODIPY–sphingosine, as described by Blom et al. (Blom et al., 2012). All cell culture reagents, except poly-D-lysine (Sigma-Aldrich), were obtained from Gibco Life Technologies (Invitrogen). Lipoprotein-deficient serum (LPDS) was prepared as described previously (Jansen et al., 2008).

Plasmid and siRNA transfections

Human PRA1 cDNA (NM_006423.1) in pCMV6-XL5 was purchased from Origene (Rockville, Maryland, USA). The LDLR-HA and HA-ubiquitin plasmids have been previously described (Hong et al., 2010; Zelcer et al., 2009). The following siRNAs were used: CTRL (Sigma-Aldrich; 5′-CGUAC-GCGGAAUACUUCGA-3′), human NDRG1 (Stealth-modified, Invitrogen; 5′-CAUCGAGACUUUACAUGGCUCUGUU-3′), mouse Ndrg1 (Qiagen FlexiTube, target sequence 5′-AACTCATTCCTGGAAACAAA-3′), human IDOL (s26522, Ambion Silencer Select, Invitrogen; 5′-GACUUUAGCCCAAUUAAUATT-3′), mouse Idol-6 (Ambion Silencer Select, Invitrogen; target sequence 5′-AA-CATTAACCTTGGTAAGAAA-3′), mouse Idol-9 (Ambion Silencer Select, Invitrogen; target sequence 5′-AGGGATCATCGAGGTTGATTA-3′), Rab4a (Qiagen validated, 5′-AAUGCAGGAACUGGCAAAUCT-3′) and PRA1 (Kim et al., 2006) (Stealth-modified, Invitrogen, 5′-GCGCCUGUUACAUUC-UCUAUCUGCG-3′). TSG101 and Hrs siRNAs have been previously described (Bache et al., 2003; Garrus et al., 2001). cDNA transfections were performed with Effectene and siRNA transfections in 10 nM final concentration with HiPerfect according to the manufacturer's instructions. To silence mouse Idol, two different siRNAs (Idol-6 and −9) were used (total 10 nM final concentration). In co-silencing experiments of two different genes, each siRNA was used at 10 nM. The siRNA transfection time was 72 hours unless otherwise stated.

Cell culture and western blot analysis

A431 cells (epidermoid carcinoma cell line, American Type Culture Collection; ATCC) and HEK293T cells (human embryonic kidney cells, ATCC) were cultured in Dulbecco's modified Eagle's medium (D-MEM) supplemented with 10% fetal bovine serum, 0.5 mM L-glutamine and 100 IU/ml penicillin, 100 µg/ml streptomycin at 37℃ in a CO_2-conditioned, humidified incubator. Oli-neu, an immortalized oligodendrocyte precursor cell line (Jung et al., 1995) was cultured in 1:1 D-MEM/HAM's F12 with B27 supplement on poly-D-lysine (50 µg/ml)-coated plastic. Cells were treated with U18666A (48 hours) in 10% FBS or loaded with LDL (18 hours) in 5% LPDS. Cells were collected for western blotting by adding 1% NP-40 including a protease inhibitor mixture (chymostatin, leupeptin, antipain and pepstatin A; 25 µg/ml each; CLAP) in PBS. The protein amount was determined by using a BioRad kit. Proteins were subjected to SDS-PAGE and western blotting as described (Jansen et al., 2008). In some experiments, the IRDye-conjugated secondary antibodies (IRDye 800CW-conjugated donkey anti-rabbit and 680LT donkey anti-mouse, LI-COR, Superior St. Lincoln, Nebraska, USA) were used prior to image capture with Odyssey Imager (LI-COR). Immunoreactive bands were quantified using ImageJ software (National Institutes of Health, Bethesda, Maryland, USA; NIH, http://rsbweb.nih.gov/ij/).

Immunofluorescence staining

Cells were fixed with 4% paraformaldehyde for 20 minutes at room temperature, and permeabilized with 0.1% or 0.3% Triton X-100 for 5 minutes. For visualization of free cholesterol, cells were stained with 0.05% filipin in 10% FBS–PBS blocking solution for 30 minutes at 37℃. The immunostainings were performed sequentially as described by Jansen et al. (Jansen et al., 2008). Images were acquired with an Olympus AX70 epifluorescence microscope (Olympus, Center Valley, Pennsylvania, USA) equipped with an Olympus DP71 CCD camera, using 20× or 40× objectives. Confocal images were acquired with a Leica TCS SP2 AOBS confocal microscope (Leica Microsystems, Wetzlar, Germany), with 40× (NA=1.25) or 60× (NA=1.4) oil objectives.

Biochemical lipid analyses

siRNA-transfected A431 cells were collected in 0.2 N NaOH and lipids were extracted as described previously (Bligh and Dyer, 1959). FC and CE were resolved on TLC plates using hexane/diethyl ether/acetic acid (80:20:1) as the mobile phase. Lipids were visualized by charring and quantified by ImageJ after scanning. To measure cholesterol synthesis, cells were pulse-labeled in serum-free medium with 20 µCi/ml [^{3}H]acetic acid in six-well plates for 2 hours at 37℃, chased for 2 hours in serum-free medium with 10 mM lovastatin and 2.5 mM mevalonate, and analyzed as described previously (Heino et al., 2000). For analysis of cholesterol efflux, [^{3}H]cholesterol (2 µCi/ml) was added to A431 cells 24 hours after siRNA transfection, the cells were incubated for an additional 8 hours, and then for 16 hours in serum-free medium in the presence of ApoA-I (10 µg/ml). The medium was collected, cells were scraped into ice-cold PBS, harvested by centrifugation and resuspended in 0.2 N NaOH. The radioactivity in medium and cells was measured by scintillation counting.

DiI–LDL binding and uptake

siRNA-transfected A431 and Oli-neu cells on coverslips were incubated with 5 µg/ml (A431 cells) or 2 µg/ml (Oli-neu cells) of DiI–LDL for 30 minutes at 4℃ (binding) or 20–30 minutes at 37℃ (uptake) in serum-free medium prior to fixation with 4% PFA and staining with DAPI for detection of the nuclei. The samples were imaged with an Olympus AX70 microscope to quantify DiI–LDL intensity. The total intensity obtained from background-substracted DiI–LDL images was divided by the number of nuclei in the image to obtain an intensity-per-cell value. Alternatively, cells were plated onto 48-well plates, transfected with siRNAs and incubated with DiI–LDL as above. Cells were washed three times with PBS prior to lysis into 200 µl of PBS–1% NP-40. The lysates were incubated on ice for 10 minutes, and 150 µl aliquots were measured in a PheraStar microplate reader (BMG Labtech, Ortenberg, Germany) with an ex540/em590 filter set. Fluorescence counts were normalized to total protein (determined from 10 µl aliquots).

Cell surface biotinylation and LDLR degradation

Cell surface biotinylation protocol was modified from Scotti et al. (Scotti et al., 2011). Cells on 60 mm plates were rinsed twice with ice-cold PBS containing 0.02 mM $CaCl_2$ and 0.15 mM $MgCl_2$ (PBS+). Cell surface proteins were biotinylated for 20 minutes with 0.5 mg/ml biotin (Sulfo-NHS-LC-Biotin; Pierce/Thermo Fisher Scientific, Rockford, Illinois, USA) diluted in PBS+. Unbound biotin was quenched by two incubations (5 minutes at 4℃) with PBS+ containing 0.1 M glycine and 0.3% BSA. The cells were washed with PBS+ and lysed (10 minutes at 4℃) in 300 µl lysis buffer (0.2% SDS, 2% NP-40 in PBS, supplemented with protease inhibitors). Lysates were collected and cleared by centrifugation, and an aliquot (20 µl) was taken for immunoblotting with the LDL receptor antibody ('total'). Biotinylated proteins from the rest of the lysate (250 µl) were precipitated with 50 µl streptavidin–agarose beads (Sigma-Aldrich)

by rotation overnight at 4℃. The beads were washed three times in cold lysis buffer and twice in PBS+. Biotinylated proteins were eluted from the beads by boiling in 2× sample buffer, separated by SDS-PAGE, and transferred to nitrocellulose for western blotting. Blots were quantified by densitometry using ImageJ. To quantify LDLR degradation, the cell surface was biotinylated and quenched as above. The cells were washed in PBS+ and chased for 0, 2, or 5 hours in complete culture medium at 37℃, lysed, and the biotinylated proteins were precipitated and LDLR amounts detected as described above.

Quantification of LDLR accumulation

Cells were transfected with control or NDRG1 siRNAs as described above and with PRA1 or pEGFP cDNA 48 hours later. At day 3, cells were fixed with 4% PFA and stained with anti-PRA1 and anti-LDLR antibodies. The number of intracellular LDLR-positive accumulations was quantified from Leica confocal images in ImageJ by counting the local maxima of the region of interest (ROI), using the 'Find maxima' plugin filter of ImageJ (http://rsbweb.nih.gov/ij/docs/menus/process.html). Cells expressing the transfected constructs (PRA1 or GFP) were defined as ROIs, and maxima were scored as 'LDLR-positive accumulations'.

Electron microscopy

Cells were cultured on glass coverslips, fixed with 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, post-fixed with 1% osmium tetroxide and 1.5% potassium ferrocyanide and dehydrated. A plastic capsule filled with embedding resin was inverted over the coverslip. After polymerization the coverslip was removed and the cells were sectioned horizontally. MVBs were systematically sampled to give a total of 60 MVBs from two individual experiments in each group, and the diameter of MVBs and number of intraluminal vesicles in the plane of sectioning were assessed in prints with a final magnification of 145000×. BSA–16 nm colloidal gold complexes were prepared according to the method of Griffiths (Griffiths, 1993). Cells were incubated with BSA–gold (OD_{520}=1) in Hepes-buffered D-MEM for 40 minutes at 37℃, fixed and processed as described above.

Ceramide quantification

Cells were washed with PBS and scraped into 800 µl of ice cold 2% NaCl. Aliquots (5 µl) were removed for protein determination. Then, 3 ml $CHCl_3$/methanol (2:1) was added, the suspension was vortexed, 1 ml of $CHCl_3$ and 1 ml of H_2O were added for phase separation, and the lower phase was collected. Lipid extracts were dried under a stream of nitrogen, lipids were re-dissolved in $CHCl_3$/methanol (2:1), and resolved with known standards on HPTLC plates with $CHCl_3$:acetic acid (9:1) as the mobile phase. HPTLC plates were dried, dipped in 3% $CuSO_4$/8% H_3PO_4, and the ceramide bands were visualized by charring with H_2SO_4 in ethanol at 180℃.

Metabolism of LDL-derived [^{3}H]SM cells was measured as previously described (Blom et al., 2012). In brief, cells were serum-starved overnight and then incubated with 50 µg/ml [^{3}H]SM/LDL for 2 hours. Cellular lipids were extracted and separated on HPTLC. Bands corresponding to lipid standards were scraped and measured by scintillation counting.

BODIPY-Cer trafficking

LDL particles were labeled with BODIPY-Cer according to (Blom et al., 2012). Cells were labeled overnight with Alexa-Fluor-568–dextran in medium containing 5% LPDS, incubated for an additional 2 hours in the absence of dextran (to chase dextran to lysosomes) and then labeled with 50 µg/ml BODIPY–Cer/LDL for 1 hours. The cells were washed and imaged by confocal microscopy at 37℃ for 4 hours. Colocalization of BODIPY-labeled ceramide and dextran was quantified using Pearson's correlation coefficient.

Immunoprecipitation of ubiquitylated proteins

Immunoprecipitation of HA-tagged proteins was performed as previously described (Zelcer et al., 2009). Briefly, equal amounts of protein in the cleared lysates were incubated with EZview^{TM} red anti-HA affinity beads (Sigma-Aldrich) for 16 hours. Subsequently, beads were washed 4× with RIPA buffer. All incubations and washes were performed at 4℃ with rotation. Proteins were eluted from the beads by boiling in 1× protein sample buffer for 5 minutes prior to western blotting and quantification.

Acknowledgements

We thank A. Uro and P. Kaipainen (University of Helsinki, Finland) for excellent technical help, and T. Commes (Université Montpellier, France), M. Carducci (Johns Hopkins University, USA), H. Stenmark (Oslo University Hospital, Norway), Peter van der Sluijs (University Medical Center Utrecht, Utrecht, Netherlands) and M. Simons (Max-Planck Institute of Experimental Medicine, Göttingen, Germany) for kindly providing the anti-NDRG1 antibody, NDRG1-Flag construct, ESCRT antibodies, Rab4 antibody and Oli-neu cells, respectively. We thank the Biomedicum Imaging Unit and the Electron Microscopy Unit of the Institute of Biotechnology, University of Helsinki, for facilities.

Author contributions

V.P., N.Z. and E.I. designed the experiments. V.P., B.V., T.B., W.W., J.N. and N.B. performed and analyzed the experiments. R.B. provided reagents. V.P., B.V., R.B., N.Z. and E.I. wrote the manuscript.

Funding

This study was supported by the Finnish Medical Foundation, Academy of Finland [grant numbers 1117064 to V.P., 131429, 131489 to E.I.]; a University of Helsinki postdoctoral researcher grant (to V.P.), the Kymenlaakso Cultural Foundation (to V.P.); the Georg and Ella Ehrnrooth Foundation; the Magnus Ehrnrooth Foundation; the Liv och Hälsa Foundation; a Career Development award from the Human Frontier Science Program Organization [grant number CDA-00057/2009 to N.Z.]; and a VIDI grant from the Netherlands Organization of Scientific Research [grant number 017.106.355 to N.Z.].

Supplementary material available online at
http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.128132/-/DC1

References

Bache, K. G., Raiborg, C., Mehlum, A. and Stenmark, H. (2003). STAM and Hrs are subunits of a multivalent ubiquitin-binding complex on early endosomes. *J. Biol. Chem.* **278**, 12513-12521.

Bandyopadhyay, S., Pai, S. K., Gross, S. C., Hirota, S., Hosobe, S., Miura, K., Saito, K., Commes, T., Hayashi, S., Watabe, M. et al. (2003). The Drg-1 gene suppresses tumor metastasis in prostate cancer. *Cancer Res.* **63**, 1731-1736.

Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., Wirkner, U., Erfle, H., Muckenthaler, M., Pepperkok, R. et al. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab.* **10**, 63-75.

Beisiegel, U., Schneider, W. J., Goldstein, J. L., Anderson, R. G. and Brown, M. S. (1981). Monoclonal antibodies to the low density lipoprotein receptor as probes for study of receptor-mediated endocytosis and the genetics of familial hypercholesterolemia. *J. Biol. Chem.* **256**, 11923-11931.

Berger, P., Sirkowski, E. E., Scherer, S. S. and Suter, U. (2004). Expression analysis of the N-Myc downstream-regulated gene 1 indicates that myelinating Schwann cells are the primary disease target in hereditary motor and sensory neuropathy-Lom. *Neurobiol. Dis.* **17**, 290-299.

Bligh, E. G. and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911-917.

Blom, T., Li, Z., Bittman, R., Somerharju, P. and Ikonen, E. (2012). Tracking sphingosine metabolism and transport in sphingolipidoses: NPC1 deficiency as a test case. *Traffic* **13**, 1234-1243.

Bucci, C., Chiariello, M., Lattero, D., Maiorano, M. and Bruni, C. B. (1999). Interaction cloning and characterization of the cDNA encoding the human prenylated rab acceptor (PRA1). *Biochem. Biophys. Res. Commun.* **258**, 657-662.

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A. et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264-278.

Carstea, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M. A., Pavan, W. J., Krizman, D. B. et al. (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* **277**, 228-231.

Echaniz-Laguna, A., Degos, B., Bonnet, C., Latour, P., Hamadouche, T., Lévy, N. and Leheup, B. (2007). NDRG1-linked Charcot-Marie-Tooth disease (CMT4D) with central nervous system involvement. *Neuromuscul. Disord.* **17**, 163-168.

Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Côté, M., Rich, R. L. et al. (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55-65.

Goldstein, J. L. and Brown, M. S. (2009). The LDL receptor. *Arterioscler. Thromb. Vasc. Biol.* **29**, 431-438.

Griffiths, G. (1993). *Fine Structure Immunohistochemistry*. Berlin: Springer-Verlag.

Guan, R. J., Ford, H. L., Fu, Y., Li, Y., Shaw, L. M. and Pardee, A. B. (2000). Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer. *Cancer Res.* **60**, 749-755.

Heino, S., Lusa, S., Somerharju, P., Ehnholm, C., Olkkonen, V. M. and Ikonen, E. (2000). Dissecting the role of the golgi complex and lipid rafts in biosynthetic transport of cholesterol to the cell surface. *Proc. Natl. Acad. Sci. USA* **97**, 8375-8380.

Hölttä-Vuori, M., Vainio, S., Kauppi, M., Van Eck, M., Jokitalo, E. and Ikonen, E. (2012). Endosomal actin remodeling by coronin-1A controls lipoprotein uptake and degradation in macrophages. *Circ. Res.* **110**, 450-455.

Hong, C., Duit, S., Jalonen, P., Out, R., Scheer, L., Sorrentino, V., Boyadjian, R., Rodenburg, K. W., Foley, E., Korhonen, L. et al. (2010). The E3 ubiquitin ligase IDOL induces the degradation of the low density lipoprotein receptor family members VLDLR and ApoER2. *J. Biol. Chem.* **285**, 19720-19726.

Hunter, M., Angelicheva, D., Tournev, I., Ingley, E., Chan, D. C., Watts, G. F., Kremensky, I. and Kalaydjieva, L. (2005). NDRG1 interacts with APO A-I and A-II and is a functional candidate for the HDL-C QTL on 8q24. *Biochem. Biophys. Res. Commun.* **332**, 982-992.

Huotari, J. and Helenius, A. (2011). Endosome maturation. *EMBO J.* **30**, 3481-3500.

Ikonen, E. (2008). Cellular cholesterol trafficking and compartmentalization. *Nat. Rev. Mol. Cell Biol.* **9**, 125-138.

Jansen, M., Pietiaïnen, V. M., Pölönen, H., Rasilainen, L., Koivusalo, M., Ruotsalainen, U., Jokitalo, E. and Ikonen, E. (2008). Cholesterol substitution increases the structural heterogeneity of caveolae. *J. Biol. Chem.* **283**, 14610-14618.

Jung, M., Krämer, E., Grzenkowski, M., Tang, K., Blakemore, W., Aguzzi, A., Khazaie, K., Chlichlia, K., von Blankenfeld, G., Kettenmann, H. et al. (1995). Lines of murine oligodendroglial precursor cells immortalized by an activated neu tyrosine kinase show distinct degrees of interaction with axons in vitro and in vivo. *Eur. J. Neurosci.* **7**, 1245-1265.

Kachhap, S. K., Faith, D., Qian, D. Z., Shabbeer, S., Galloway, N. L., Pili, R., Denmeade, S. R., DeMarzo, A. M. and Carducci, M. A. (2007). The N-Myc down regulated Gene1 (NDRG1) Is a Rab4a effector involved in vesicular recycling of E-cadherin. *PLoS ONE* **2**, e844.

Kalaydjieva, L., Hallmayer, J., Chandler, D., Savov, A., Nikolova, A., Angelicheva, D., King, R. H., Ishpekova, B., Honeyman, K., Calafell, F. et al. (1996). Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q24. *Nat. Genet.* **14**, 214-217.

Kalaydjieva, L., Nikolova, A., Turnev, I., Petrova, J., Hristova, A., Ishpekova, B., Petkova, I., Shmarov, A., Stancheva, S., Middleton, L. et al. (1998). Hereditary motor and sensory neuropathy—Lom, a novel demyelinating neuropathy associated with deafness in gypsies. Clinical, electrophysiological and nerve biopsy findings. *Brain* **121**, 399-408.

Kalaydjieva, L., Gresham, D., Gooding, R., Heather, L., Baas, F., de Jonge, R., Blechschmidt, K., Angelicheva, D., Chandler, D., Worsley, P. et al. (2000). N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. *Am. J. Hum. Genet.* **67**, 47-58.

Kim, J. T., Cho, M. Y., Choi, S. C., Kim, J. W., Chae, S. K., Yoon, D. Y., Kim, J. W. and Lim, J. S. (2006). Prenylated Rab acceptor 1 (PRA1) inhibits TCF/beta-catenin signaling by binding to beta-catenin. *Biochem. Biophys. Res. Commun.* **349**, 200-208.

Kim, J. T., Kim, J. W., Kang, Y. H., Kim, K. D., Lee, S. J., Choi, S. C., Kim, K. S., Chae, S. K., Kim, J. W., Lim, J. S. et al. (2012). NDRG2 and PRA1 interact and synergistically inhibit T-cell factor/β-catenin signaling. *FEBS Lett.* **586**, 3962-3968.

King, R. H., Chandler, D., Lopaticki, S., Huang, D., Blake, J., Muddle, J. R., Kilpatrick, T., Nourallah, M., Miyata, T., Okuda, T. et al. (2011). Ndrg1 in development and maintenance of the myelin sheath. *Neurobiol. Dis.* **42**, 368-380.

Ko, D. C., Gordon, M. D., Jin, J. Y. and Scott, M. P. (2001). Dynamic movements of organelles containing Niemann-Pick C1 protein: NPC1 involvement in late endocytic events. *Mol. Biol. Cell* **12**, 601-614.

Kobuna, H., Inoue, T., Shibata, M., Gengyo-Ando, K., Yamamoto, A., Mitani, S. and Arai, H. (2010). Multivesicular body formation requires OSBP-related proteins and cholesterol. *PLoS Genet.* **6**.

Lachat, P., Shaw, P., Gebhard, S., van Belzen, N., Chaubert, P. and Bosman, F. T. (2002). Expression of NDRG1, a differentiation-related gene, in human tissues. *Histochem. Cell Biol.* **118**, 399-408.

Lakadamyali, M., Rust, M. J. and Zhuang, X. (2006). Ligands for clathrin-mediated endocytosis are differentially sorted into distinct populations of early endosomes. *Cell* **124**, 997-1009.

Lalli, G. and Hall, A. (2005). Ral GTPases regulate neurite branching through GAP-43 and the exocyst complex. *J. Cell Biol.* **171**, 857-869.

Lee, S. M., Chin, L. S. and Li, L. (2012). Charcot-Marie-Tooth disease-linked protein SIMPLE functions with the ESCRT machinery in endosomal trafficking. *J. Cell Biol.* **199**, 799-816.

Ligon, K. L., Fancy, S. P., Franklin, R. J. and Rowitch, D. H. (2006). Olig gene function in CNS development and disease. *Glia* **54**, 1-10.

Liu, H. P., Wu, C. C., Kao, H. Y., Huang, Y. C., Liang, Y., Chen, C. C., Yu, J. S. and Chang, Y. S. (2011). Proteome-wide dysregulation by PRA1 depletion delineates a role of PRA1 in lipid transport and cell migration. *Mol. Cell Proteomics* **10**, M900641-MCP200.

Loftus, S. K., Morris, J. A., Carstea, E. D., Gu, J. Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M. A., Tagle, D. A. et al. (1997). Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* **277**, 232-235.

MacDonald, C., Buchkovich, N. J., Stringer, D. K., Emr, S. D. and Piper, R. C. (2012). Cargo ubiquitination is essential for multivesicular body intralumenal vesicle formation. *EMBO Rep.* **13**, 331-338.

Miaczynska, M., Pelkmans, L. and Zerial, M. (2004). Not just a sink: endosomes in control of signal transduction. *Curr. Opin. Cell Biol.* **16**, 400-406.

Mukhopadhyay, D. and Riezman, H. (2007). Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201-205.

Nelissen, K., Mulder, M., Smets, I., Timmermans, S., Smeets, K., Ameloot, M. and Hendriks, J. J. (2012). Liver X receptors regulate cholesterol homeostasis in oligodendrocytes. *J. Neurosci. Res.* **90**, 60-71.

Okuda, T., Higashi, Y., Kokame, K., Tanaka, C., Kondoh, H. and Miyata, T. (2004). Ndrg1-deficient mice exhibit a progressive demyelinating disorder of peripheral nerves. *Mol. Cell. Biol.* **24**, 3949-3956.

Okuda, T., Kokame, K. and Miyata, T. (2008). Differential expression patterns of NDRG family proteins in the central nervous system. *J. Histochem. Cytochem.* **56**, 175-182.

Piper, R. C. and Katzmann, D. J. (2007). Biogenesis and function of multivesicular bodies. *Annu. Rev. Cell Dev. Biol.* **23**, 519-547.

Piquemal, D., Joulia, D., Balaguer, P., Basset, A., Marti, J. and Commes, T. (1999). Differential expression of the RTP/Drg1/Ndr1 gene product in proliferating and growth arrested cells. *Biochim. Biophys. Acta* **1450**, 364-373.

Raiborg, C. and Stenmark, H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* **458**, 445-452.

Reddy, J. V., Ganley, I. G. and Pfeffer, S. R. (2006). Clues to neuro-degeneration in Niemann-Pick type C disease from global gene expression profiling. *PLoS ONE* **1**, e19.

Roberts, R. C., Peden, A. A., Buss, F., Bright, N. A., Latouche, M., Reilly, M. M., Kendrick-Jones, J. and Luzio, J. P. (2010). Mistargeting of SH3TC2 away from the recycling endosome causes Charcot-Marie-Tooth disease type 4C. *Hum. Mol. Genet.* **19**, 1009-1018.

Saher, G., Brügger, B., Lappe-Siefke, C., Möbius, W., Tozawa, R., Wehr, M. C., Wieland, F., Ishibashi, S. and Nave, K. A. (2005). High cholesterol level is essential for myelin membrane growth. *Nat. Neurosci.* **8**, 468-475.

Scotti, E., Hong, C., Yoshinaga, Y., Tu, Y., Hu, Y., Zelcer, N., Boyadjian, R., de Jong, P. J., Young, S. G., Fong, L. G. et al. (2011). Targeted disruption of the idol gene alters cellular regulation of the low-density lipoprotein receptor by sterols and liver x receptor agonists. *Mol. Cell. Biol.* **31**, 1885-1893.

Sidiropoulos, P. N., Miehe, M., Bock, T., Tinelli, E., Oertli, C. I., Kuner, R., Meijer, D., Wollscheid, B., Niemann, A. and Suter, U. (2012). Dynamin 2 mutations in Charcot-Marie-Tooth neuropathy highlight the importance of clathrin-mediated endocytosis in myelination. *Brain* **135**, 1395-1411.

Sivars, U., Aivazian, D. and Pfeffer, S. R. (2003). Yip3 catalyses the dissociation of endosomal Rab-GDI complexes. *Nature* **425**, 856-859.

Sorrentino, V. and Zelcer, N. (2012). Post-transcriptional regulation of lipoprotein receptors by the E3-ubiquitin ligase inducible degrader of the low-density lipoprotein receptor. *Curr. Opin. Lipidol.* **23**, 213-219.

Staub, O. and Rotin, D. (2006). Role of ubiquitylation in cellular membrane transport. *Physiol. Rev.* **86**, 669-707.

Stendel, C., Roos, A., Kleine, H., Arnaud, E., Ozçelik, M., Sidiropoulos, P. N., Zenker, J., Schüpfer, F., Lehmann, U., Sobota, R. M. et al. (2010). SH3TC2, a protein mutant in Charcot-Marie-Tooth neuropathy, links peripheral nerve myelination to endosomal recycling. *Brain* **133**, 2462-2474.

Stuffers, S., Sem Wegner, C., Stenmark, H. and Brech, A. (2009). Multivesicular endosome biogenesis in the absence of ESCRTs. *Traffic* **10**, 925-937.

Sugii, S., Reid, P. C., Ohgami, N., Du, H. and Chang, T. Y. (2003). Distinct endosomal compartments in early trafficking of low density lipoprotein-derived cholesterol. *J. Biol. Chem.* **278**, 27180-27189.

Trajkovic, K., Hsu, C., Chiantia, S., Rajendran, L., Wenzel, D., Wieland, F., Schwille, P., Brügger, B. and Simons, M. (2008). Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244-1247.

van Belzen, N., Dinjens, W. N., Diesveld, M. P., Groen, N. A., van der Made, A. C., Nozawa, Y., Vlietstra, R., Trapman, J. and Bosman, F. T. (1997). A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms. *Lab. Invest.* **77**, 85-92.

van der Goot, F. G. and Gruenberg, J. (2006). Intra-endosomal membrane traffic. *Trends Cell Biol.* **16**, 514-521.

Verhoeven, K., De Jonghe, P., Coen, K., Verpoorten, N., Auer-Grumbach, M., Kwon, J. M., FitzPatrick, D., Schmedding, E., De Vriendt, E., Jacobs, A. et al. (2003). Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot-Marie-Tooth type 2B neuropathy. *Am. J. Hum. Genet.* **72**, 722-727.

Wang, Y., Huang, Y., Hobbs, H. H. and Cohen, J. C. (2012). Molecular characterization of proprotein convertase subtilisin/kexin type 9-mediated degradation of the LDLR. *J. Lipid Res.* **53**, 1932-1943.

Wicher, G. and Aldskogius, H. (2008). Megalin deficiency induces critical changes in mouse spinal cord development. *Neuroreport* **19**, 559-563.

Zelcer, N., Hong, C., Boyadjian, R. and Tontonoz, P. (2009). LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor. *Science* **325**, 100-104.

Zhang, L., Reue, K., Fong, L. G., Young, S. G. and Tontonoz, P. (2012). Feedback regulation of cholesterol uptake by the LXR-IDOL-LDLR axis. *Arterioscler. Thromb. Vasc. Biol.* **32**, 2541-2546.

Zhao, S., Hu, X., Park, J., Zhu, Y., Zhu, Q., Li, H., Luo, C., Han, R., Cooper, N. and Qiu, M. (2007). Selective expression of LDLR and VLDLR in myelinating oligodendrocytes. *Dev. Dyn.* **236**, 2708-2712.

The American Journal of Pathology, Vol. 185, No. 4, April 2015

The American Journal of
PATHOLOGY
ajp.amjpathol.org

EPITHELIAL AND MESENCHYMAL CELL BIOLOGY

Elevated Levels of StAR-Related Lipid Transfer Protein 3 Alter Cholesterol Balance and Adhesiveness of Breast Cancer Cells

Potential Mechanisms Contributing to Progression of HER2-Positive Breast Cancers

Boris Vassilev,[*] Harri Sihto,[†] Shiqian Li,[*‡] Maarit Hölttä-Vuori,[*‡] Jaakko Ilola,[*] Johan Lundin,[§] Jorma Isola,[¶] Pirkko-Liisa Kellokumpu-Lehtinen,[¶] Heikki Joensuu,[‖] and Elina Ikonen[*‡]

From the Faculty of Medicine, Department of Anatomy, and the Laboratory of Molecular Oncology,[†] Translational Cancer Biology Program, University of Helsinki, Helsinki; the Minerva Foundation Institute for Medical Research,[‡] Biomedicum 2U, Helsinki; the Institute for Molecular Medicine Finland,[§] University of Helsinki, Biomedicum Helsinki 2U, Helsinki; the Institute of Medical Technology,[¶] University of Tampere and Tampere University Central Hospital, Tampere; and the Department of Oncology,[‖] Helsinki University Central Hospital, Helsinki, Finland*

Accepted for publication December 9, 2014.

Address correspondence to Elina Ikonen, M.D., Ph.D., Faculty of Medicine, Department of Anatomy, University of Helsinki, Haartmaninkatu 8, FI-00290 Helsinki, Finland. E-mail: elina.ikonen@helsinki.fi.

The *STARD3* gene belongs to the minimal amplicon in *HER2*-positive breast cancers and encodes a cholesterol-binding membrane protein. To study how elevated StAR-related lipid transfer protein 3 (StARD3) expression affects breast cancer cells, we generated MCF-7 cells stably overexpressing StARD3—green fluorescent protein. We found that StARD3-overexpressing cells exhibited non-adherent morphological features, had increased Src levels, and had altered cholesterol balance, as evidenced by elevated mRNA levels of the cholesterol biosynthesis rate-limiting enzyme 3-hydroxy-3-methylglutaryl-coenzyme A reductase, and increased plasma membrane cholesterol content. On removal of serum and insulin from the culture medium, the morphological characteristics of the StARD3-overexpressing cells changed, the cells became adherent, and they developed enlarged focal adhesions. Under these conditions, the StARD3-overexpressing cells maintained elevated Src and plasma membrane cholesterol content and showed increased phosphorylation of focal adhesion kinase. In two Finnish nationwide patient cohorts, approximately 10% (212/2220) breast cancers exhibited high StARD3 protein levels, which was strongly associated with *HER2* amplification; several factors related to poor disease outcome and poor breast cancer—specific survival. In addition, high StARD3 levels in breast cancers were associated with elevated 3-hydroxy-3-methylglutaryl-coenzyme A reductase mRNA levels and anti—Src-Tyr416 immunoreactivity. These results provide evidence that StARD3 overexpression results in increased cholesterol biosynthesis and Src kinase activity in breast cancer cells and suggest that elevated StARD3 expression may contribute to breast cancer aggressiveness by increasing membrane cholesterol and enhancing oncogenic signaling. *(Am J Pathol 2015, 185: 987—1000; http://dx.doi.org/10.1016/j.ajpath.2014.12.018)*

The genomic region in chromosome 17 containing the *HER2* (*ERBB2*) oncogene (17q12-q21) is frequently amplified in cancers. This region contains multiple genes coamplified with *HER2*.[1,2] Several of these genes, including *STARD3*, have been suggested to functionally contribute to the proliferation of

Supported by the Helsinki Biomedical Graduate Programme (B.V.), Academy of Finland grants 131489, 263841, and 272130 (E.I.), Institute for Molecular Medicine Finland national network funding (E.I.), and Sigrid Juselius Foundation (E.I.).

Disclosures: None declared.

http://dx.doi.org/10.1016/j.ajpath.2014.12.018

HER2-positive cells, on the basis of RNA interference studies.[3,4] The receptor tyrosine-protein kinase erbB-2 (ErbB2) protein is a cell surface transmembrane tyrosine kinase that forms long-lived, potent heterodimers with other receptors of the epidermal growth factor receptor family, and acts as a networking receptor that mediates growth, survival, migration, and proliferation signals to cancer cells.[5,6] A monoclonal antibody directed against the extracellular domain of ErbB2, trastuzumab, forms the basis for the treatment of *HER2*-positive breast cancers. However, more than half of *HER2*-positive cancers either are or become resistant to the drug, implying that other genes, such as those coamplified with *HER2*, may contribute to the aggressive behavior of these tumors.[3,4,7]

The StAR-related lipid transfer protein 3 (StARD3; alias MLN64 for metastatic lymph node 64) transcript was originally identified as being highly expressed in 14 of 93 invasive breast carcinomas that also expressed high *HER2* mRNA levels.[8] The *STARD3* transcript encodes a ubiquitously expressed protein that localizes to the late endosomal organelles (LEs)[9] and binds cholesterol.[10] Its N-terminal domain with four transmembrane spans targets the protein to the LE membrane,[9,11] and its C-terminal START domain harbors a hydrophobic cholesterol-binding pocket that projects toward the cytoplasm.[10] The N-terminal domain is also capable of interacting with cholesterol.[12] Despite considerable interest in the StARD3 protein, its precise function is not well understood. Targeted disruption of the START domain causes only modest alterations in whole body cholesterol metabolism in mice.[13] However, there are several indications that cholesterol binding or transfer by StARD3 affects cellular functions. The cholesterol-binding activity of StARD3 is involved in controlling the association of LEs with the actin cytoskeleton and the actin-dependent movement of LEs.[14] The protein can also facilitate the transfer of cholesterol between LEs and other membrane compartments, such as mitochondria[15] and the plasma membrane.[11] The nascent StARD3 protein is transported to the LEs via the plasma membrane.[16]

Recent work suggests that the coamplification of *STARD3* with *ERBB2* might play a role in breast cancer progression. First, StARD3 overexpression, as measured by quantitative real-time PCR, increased the prognostic power of ErbB2 overexpression for disease-free survival of breast cancer patients.[17] Second, knockdown of StARD3 by siRNA in *HER2*-positive breast cancer cell lines (both trastuzumab-responding cell lines BT474 and SKBR3 and non-responding lines JIMT1 and KPL4) led to induction of apoptosis, suggesting that StARD3 is required for the growth and survival of these cells.[4] However, the mechanisms by which StARD3 may mediate these effects are not known.

To learn whether increased StARD3 protein amounts can alter the characteristics of cancer cells independently of *HER2* amplification, we investigated the effects of StARD3 overexpression in a well-studied model of *HER2*-negative human breast cancer MCF-7 cells. We provide evidence that the elevated expression of StARD3 protein is accompanied by significant morphological and biochemical alterations in these cells, including altered membrane cholesterol content and enhanced signaling related to cell adhesion. We also raised a polyclonal StARD3-specific antibody and demonstrate high StARD3 immunoreactivity, coinciding with *HER2* amplification, in 1 of 10 human breast tumors in two cohorts of >2000 patients. High StARD3 levels also associate with indicators of altered cholesterol metabolism and signaling in these tumors. Together, our data suggest that, in highly StARD3-expressing breast cancer cells, the cholesterol-mobilizing and membrane-altering properties of the protein might exaggerate oncogenic signaling via membrane-associated kinases and thereby contribute to breast cancer aggressiveness.

Materials and Methods

Antibodies

To raise anti-StARD3 antibodies, the START domain of the StARD3 protein (corresponding to amino acids 230 to 445) was expressed as a His6−glutathione *S*-transferase fusion protein in *Escherichia coli* JM109(DE3) and purified from cell debris on preparative SDS-polyacrylamide gels. New Zealand white rabbits were immunized with the purified protein. The polyclonal antiserum was affinity purified using the glutathione *S*-transferase−fusion protein and used at 1:10 dilution for immunofluorescence staining and at 1:50 to 1:200 dilution for immunoblotting. Anti-ErbB2 mouse monoclonal antibody (3B5; ab16901) was purchased from Abcam (Cambridge, UK), Lamp-1 H43A-c mouse monoclonal antibody was from the Developmental Studies Hybridoma Bank (University of Iowa, Iowa City), sterol regulatory element-binding protein 2 mouse monoclonal antibody (557037) was from BD Biosciences (San Jose, CA), phospho-Src (Tyr416) rabbit polyclonal antibody (number 2101) was from Cell Signaling Technology (Danvers, MA), anti-focal adhesion kinase (pp125FAK) rabbit antibody (F2918) was from Sigma-Aldrich (St. Louis, MO), and FAK (pY397) purified mouse antibody (611807) was from BD Transduction Laboratories (San Jose, CA).

Cell Culture

MCF-7 cells were cultured in Eagle's minimum essential medium from Lonza (Verviers, Belgium), supplemented with 10% fetal bovine serum (10270-106) from Life Technologies (Grand Island, NY), 0.5 mmol/L L-glutamine (25030-024) from Life Technologies, 0.01 mg/mL insulin (human, rDNA) from Novo Nordisk (Bagsværd, Denmark), MEM nonessential amino acids (M7145) from Sigma-Aldrich, and 100 IU/mL penicillin and 100 μg/mL streptomycin (DE17-602E) from Lonza. Flasks and dishes were purchased from Nunc (Roskilde, Denmark), and 12- and 6-well dishes were from Falcon (Becton-Dickinson Labware, Franklin Lakes, NJ).

For the control cell line, a pEGFP-C1 empty plasmid from BD Biosciences Clontech (Palo Alto, CA) was used. The plasmid encoding green fluorescent protein (GFP)−fused full-length

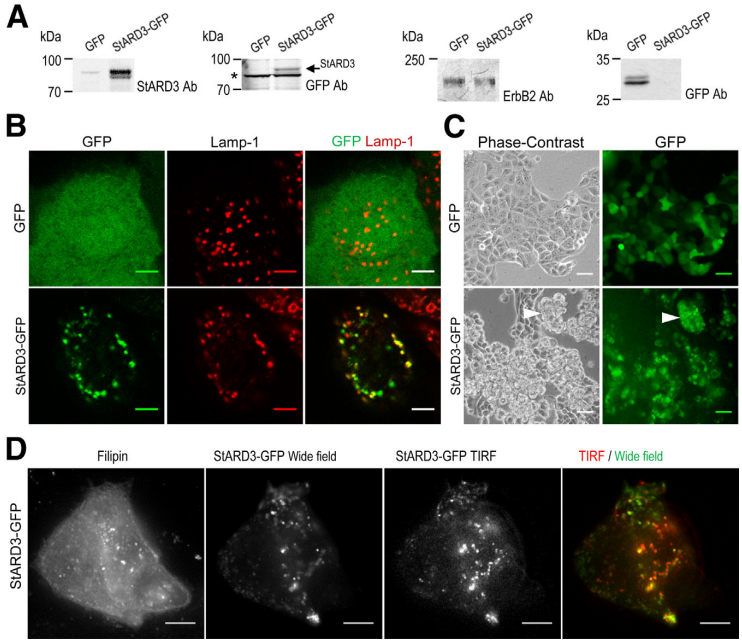


Figure 1 StAR-related lipid transfer protein 3—green fluorescent protein (StARD3-GFP) overexpression in breast cancer MCF-7 cells results in clustered growth of rounded-up cells without increasing ErbB2 levels. **A:** Immunoblots of cell lysates from control GFP and StARD3-GFP—expressing MCF-7 cell lines. Equal amounts of total protein were separated by SDS-PAGE, transferred by Western blot analysis, and detected by immunostaining with the indicated antibodies (Abs). The anti-GFP antibody detects a cross-reactive band (**asterisk**) close to the StARD3 band (**arrow**). **B:** Fluorescence micrographs of GFP and StARD3-GFP cells. Cells were fixed with paraformaldehyde, immunostained with anti—Lamp-1 antibodies, and imaged with a confocal microscope at the focal plane with most prominent Lamp-1 staining. **C:** Micrographs of live cells. Phase-contrast and wide-field fluorescence images. **Arrowheads** indicate a strongly StARD3-GFP—expressing, poorly adherent cell cluster. **D:** Wide-field fluorescence and total internal reflection fluorescence (TIRF) micrograph of a StARD3-GFP—expressing cell. Cells were fixed with PFA, stained with filipin, and imaged with a TIRF microscope for plasma membrane visualization. Scale bars: 5 μm (**B**); 20 μm (**C**); 10 μm (**D**).

StARD3 has been described.[14] To establish a StARD3-overexpressing cell line, cells were seeded at 1.25×10^5 cells per well on 24-well plates and transfected with 1 μg plasmid with Lipofectamine-LTX (Life Technologies) at 1:3 ratio, according to the manufacturer's protocol. After 48 hours, cells were passaged into medium containing 0.4 or 0.75 mg/mL geneticin G-418 sulfate (Life Technologies) and selected for 3 weeks before single-cell cloning. Expression efficiency of single clones was screened by fluorescence microscopy. For the StARD3-GFP—overexpressing cell line, single-cell cloning was performed a second time to obtain higher StARD3-GFP expression. Both stable cell lines were maintained with 0.2 mg/mL geneticin G-418 sulfate. The GFP control cell line was split twice a week at 1:3 and 1:6 ratios. The StARD3-GFP cell line was split once a week at 1:3 and 1:6 ratios, and medium was changed once between splitting. To inhibit cholesterol synthesis, cells were treated for 24 hours with 10 μmol/L lovastatin (Merck & Co, White House Station, NJ) dissolved in dimethyl sulfoxide (Sigma-Aldrich). Control samples were treated with dimethyl sulfoxide only.

Immunoblotting

Cell lysates for Western blot analysis were collected in denaturing lysis buffer containing 5 mmol/L EDTA and 1% SDS in phosphate-buffered saline, with added protease inhibitors [chymostatin, leupeptin, antipain, and pepstatin A (all from Sigma-Aldrich) at a final concentration of 25 μg/μL each] and phosphatase inhibitors [activated Na-ortho-vanadate (Sigma-Aldrich) at 2 mmol/L final concentration and NaF (Sigma-Aldrich) at 25 mmol/L final concentration]. For Western blot analysis, comparable amounts of cells were collected in denaturing lysis buffer containing protease inhibitors and chymostatin, leupeptin, antipain, and pepstatin A. Protein concentration was determined with the DC Protein Assay from Bio-Rad Laboratories (Hercules, CA). Proteins (20 μg of protein per lane) were separated by SDS-PAGE, transferred to a nitrocellulose membrane, immunoblotted and detected with an Odyssey infrared scanner from LI-COR (Lincoln, NE) or, alternatively, with Amersham ECL Western Blotting Detection Reagents (RPN2106) on Amersham Hyperfilm ECL (28-9068-36) from GE Healthcare Limited (Buckinghamshire, UK). The resulting images were quantified with ImageJ software version 1.48 (NIH, Bethesda, MD).[18]

Fluorescence Staining and Imaging of Cells

For imaging focal adhesions (FAs), cells were fixed and permeabilized with methanol for 5 minutes at −20°C. Otherwise, cells on coverslips were fixed with 4% paraformaldehyde for 20 minutes at room temperature, and autofluorescence was quenched in 50 mmol/L ammonium chloride for 15 minutes at room temperature. For filipin staining and permeabilization, the cells were fixed and stained with 0.05% filipin (F9765) from Sigma-Aldrich in 10% fetal bovine serum in phosphate-buffered saline for 30 minutes at 37°C. Otherwise, cells were permeabilized with 0.1% Triton X-100 (T8787) from Sigma-Aldrich in phosphate-buffered saline for 5 minutes at room temperature and blocked with 10% fetal bovine serum for 30 minutes at 37°C. Primary and secondary antibody incubation was performed sequentially. Red/green fluorescence images

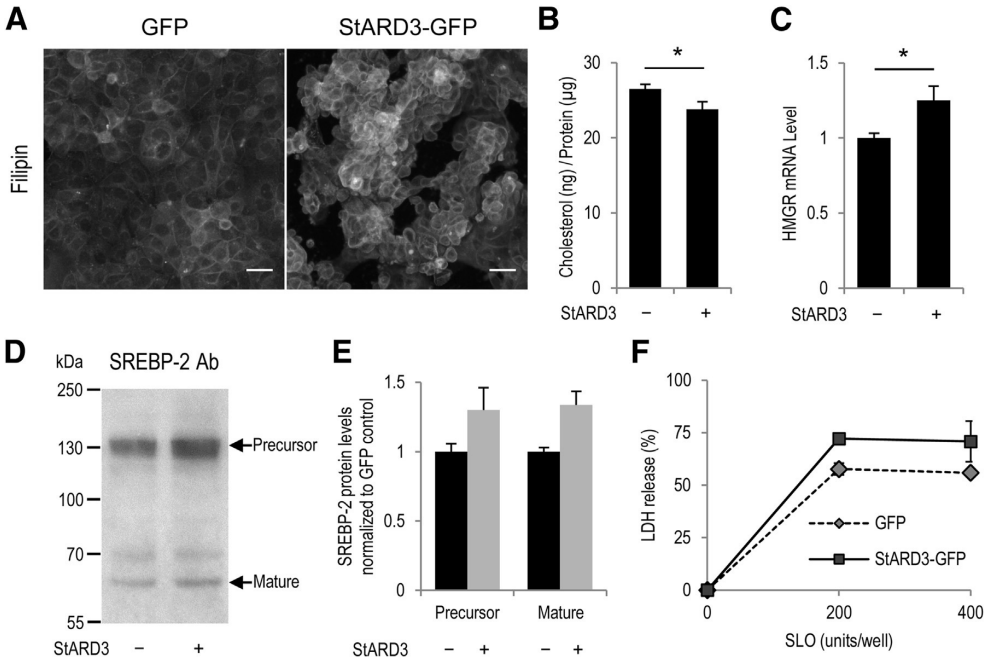


Figure 2 StAR-related lipid transfer protein 3—green fluorescent protein (StARD3-GFP) overexpression results in increased plasma membrane cholesterol and 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) transcript levels. **A:** Filipin staining. Cells grown in full medium were fixed with paraformaldehye, stained with filipin, and imaged with a wide-field fluorescence microscope. **B:** Determination of cholesterol levels. Lipids were extracted, and the amount of free cholesterol was analyzed biochemically. **C:** Analysis of HMGR transcript levels. Total RNA extracted from cells was reverse transcribed, and HMGR cDNA was quantified by real-time RT-PCR. **D:** Immunoblots of cell lysates. Equal protein amounts were separated by SDS-PAGE, transferred by Western blot analysis, and detected by immunostaining with sterol regulatory element-binding protein 2 (SREBP-2) antibody. The approximately 130-kDa precursor and the approximately 66-kDa mature form are indicated. **E:** Quantification of SREBP-2 precursor and mature form levels by densitometry. $P = 0.007$ versus GFP control cell line for SREBP-2 mature form. **F:** Sensitivity of cells to streptolysin O (SLO). Resistance of the plasma membrane to SLO was measured as the fraction of total cellular lactate dehydrogenase (LDH) released to the medium after 30 minutes of treatment at 37°C. The measurements at 0 U per well SLO were lower than the sensitivity of the assay and set to 0. Data represent means ± SEM for content of cholesterol (ng) per protein (µg) (**B**); HMGR transcript levels normalized to control (MCF-7—GFP cells) (**C**); seven replicates per condition from two independent experiments normalized to control GFP cells (**E**); two independent experiments (**F**). $n = 6$ from two independent experiments (**B**); $n = 7$ from three independent experiments (**C**). *$P < 0.05$ versus GFP control cell line (Student's *t*-test). Scale bar = 20 µm (**A**).

were acquired with an Olympus AX70 microscope from Olympus Corporation (Hamburg, Germany) equipped with an Olympus DP71 charge-coupled device camera using 10× or 20× air objectives. Filipin (UV) images were acquired with the same microscope and with an inverted Olympus microscope with a Polychrome IV monochromator by TILL Photonics (Gräfelfing, Germany) equipped with a TILL IMAGO-QE camera using a 10× air objective. Confocal images were acquired with a Leica TCS SP2 AOBS confocal microscope from Leica Microsystems (Mannheim, Germany) with a 63× oil immersion objective. Total internal reflection fluorescence microscopy images were acquired with a Nikon Eclipse Ti-E inverted microscope (Nikon Corporation, Kanagawa, Japan) equipped with an argon laser and a 100× oil immersion objective (numerical aperture, 1.49). The corresponding wide-field images of filipin and StARD3-GFP fluorescence were taken at a focal plane 1 µm above the total internal reflection fluorescence. Phase-contrast and green fluorescent images of live cells and fixed cells were acquired with a Zeiss AxioVert 200 microscope from Carl Zeiss Jena (Jena, Germany) equipped with a Zeiss AxioCam HRc color camera using 10× air objective. All quantifications were done with ImageJ software version 1.48.[18]

Lipid Extraction and Quantification

Lipids were extracted using the Bligh and Dyer method.[19] Free cholesterol was resolved on a high performance thin layer chromatography plate using hexane/diethyl ether/acetic acid (80:20:1) as the mobile phase, visualized by charring, and quantified as described previously.[20]

SLO Permeabilization

Streptolysin O (SLO) from *Streptococcus pyogenes* (S0149), β-nicotinamide adenine dinucleotide, reduced dipotassium salt (N4505), and sodium pyruvate (P2256) were purchased from Sigma-Aldrich. SLO was dissolved in deionized water to 50 U/µL, and stored at −70°C. SLO activation and membrane permeabilization were performed

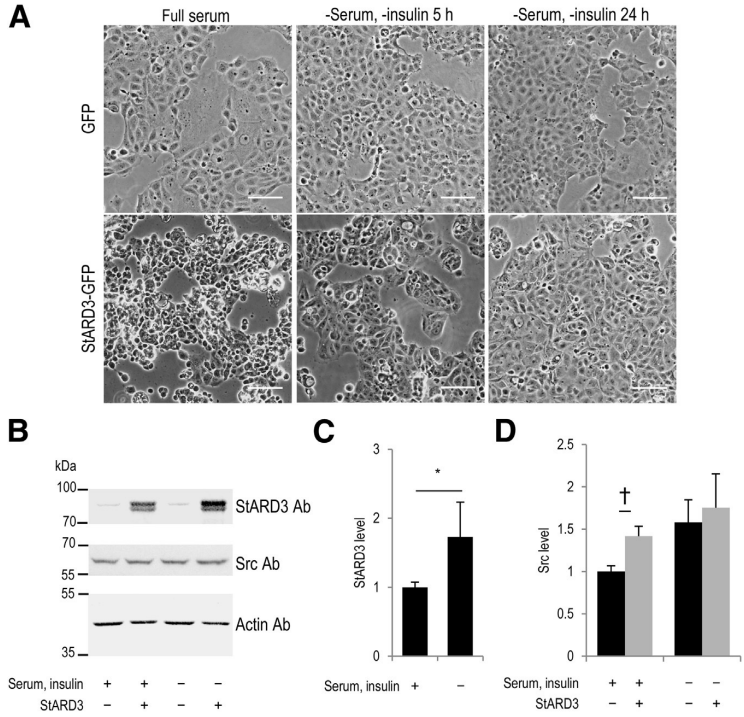


Figure 3 Removal of serum and insulin from the growth medium increases the adherence of StAR-related lipid transfer protein 3 (StARD3)–overexpressing cells. **A:** Micrographs of cells grown in the presence or absence of serum and insulin. Phase-contrast images of green fluorescent protein (GFP) and StARD3-GFP cells deprived of serum and insulin for 0, 5, or 24 hours and fixed with paraformaldehyde. **B:** Immunoblots of cell lysates. Representative blots of GFP (StARD3-) and StARD3-GFP (StARD3+) cells grown in full medium (+serum, +insulin) or deprived of serum and insulin (−serum, −insulin) for 24 hours. Equal amounts of total protein were separated by SDS-PAGE, transferred by Western blot analysis, and detected with anti-StARD3, anti-Src, or anti-actin antibodies (Abs). **C:** Quantification of StARD3 levels by densitometry. **D:** Quantification of Src levels by densitometry. Data represent means ± SEM for six replicates per condition from three independent experiments normalized to control (GFP cells, +serum, +insulin) (**C**) or six replicates per condition from three independent experiments normalized to control (GFP control cells, +serum, +insulin) (**D**). $^{*}P < 0.05$ cells in −serum, −insulin versus complete medium (Student's *t*-test); $^{\dagger}P < 0.05$ versus GFP control cell line (Student's *t*-test). Scale bar = 40 μm.

as previously described.[21] Briefly, nonconfluent cells in 12-well plates were washed and incubated in potassium acetate buffer [115 mmol/L potassium acetate (Sigma-Aldrich) and 25 mmol/L HEPES (Amresco, Solon, OH)] for 30 minutes at 37°C. SLO was activated by adding dithiothreitol to a final concentration of 10 mmol/L and incubating for 30 minutes at 37°C. The activated SLO was immediately added to the cells at the indicated amounts and incubated for 30 minutes at 37°C. The medium was collected, and the cells were lysed with 0.1% Triton X-100 (Sigma-Aldrich) in potassium acetate buffer. The decrease in absorbance at 340 nm, resulting from the oxidation of NADH on addition of pyruvate to medium and cell lysate, was measured over 1 minute. The fraction of lactate dehydrogenase (LDH) released to the medium after membrane permeabilization was calculated as the ratio of LDH in the medium/sum of LDH in the medium and in the cell lysate in each sample.

Patient Samples

A Finnish nationwide population-based breast cancer series of primary tumors, FinProg, was collected as described previously.[22] In brief, women diagnosed with breast cancer in five well-defined areas of Finland from 1991 to 1992 were identified, and individual clinicopathological data were collected from the hospital case records and tumor tissue samples from the archives of the pathological departments. Those patients whose StARD3 staining was noninformative were excluded from the study. In total, 1325 of 2930 patients were included in the study. In another study, FinHer,[23] axillary node–positive or high-risk node–negative samples were collected from 2000 to 2003 from women who had undergone a breast surgery with auxiliary node dissection or sentinel node biopsy for invasive breast carcinoma. Herein, we included a total of 895 patients of the 1010 who participated in the study (approximately 40% of the eligible women in Finland who received a diagnosis of breast cancer within this period). Permission to use formalin-fixed, paraffin-embedded tissues for research purposes from the FinProg study was provided as per local legislation by the Ministry of Social Affairs and Health, Finland (permission 123/08/97). An ethics committee at the Helsinki University Central Hospital (Helsinki, Finland) approved the FinHer study and the current study (permission HUS 125/13/03/02/2014). Study participants provided written informed consent before tissue sample collection or initiation of the study treatments.

Tissue microarrays were constructed as described earlier.[22] Tissue microarray sections (5 μm thick) were cut on SuperFrost+ slides from Menzel-Gläser (Braunschweig, Germany), deparaffinized in xylene, and rehydrated through a decreasing alcohol gradient. Endogenous peroxidase activity was blocked with 1% hydrogen peroxide. Antigen retrieval for Src-Tyr416 was performed in 10 mmol/L sodium citrate (pH 6.0) using an autoclave at 120°C for 2 minutes. Antigen retrieval for StARD3 was performed in Tris-EDTA (10 mmol/L Tris-Base and 1 mmol/L EDTA, pH 9.0) in a water bath at 98°C for 20

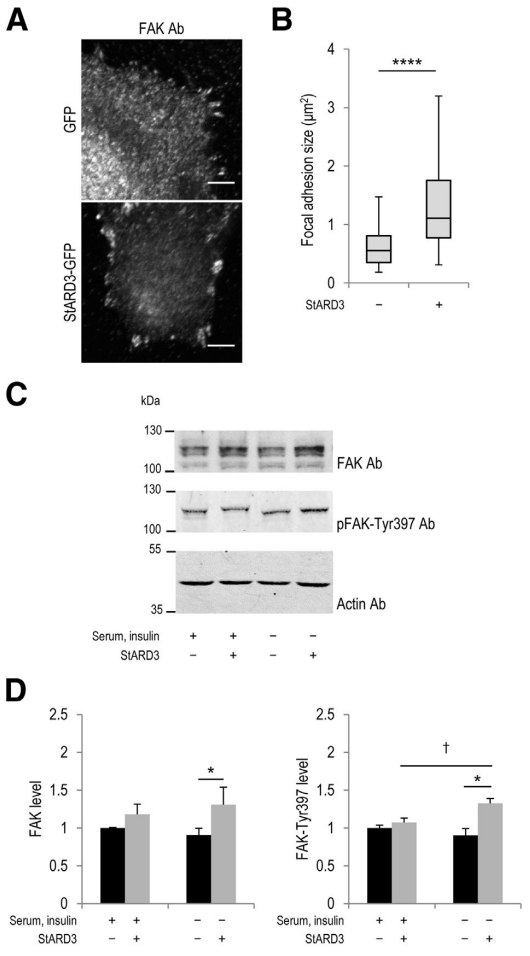


Figure 4 StAR-related lipid transfer protein 3 (StARD3) overexpression causes an increase in focal adhesion size and increased phosphorylation of focal adhesion kinase (FAK) in the absence of serum and insulin. **A:** FAK staining in the absence of serum and insulin. Cells were deprived of serum and insulin for 24 hours, fixed with methanol, immunostained with anti-FAK antibodies (Abs), and imaged with a confocal microscope at the basal plane, which has the most prominent FAK staining. **B:** Quantification of focal adhesion size from fluorescence micrographs. The size of FAK-stained patches was analyzed from micrographs (10 cells for each condition). Box plot shows focal adhesion size in μm^2. Green fluorescent protein (GFP) cells: median = 0.55, n = 83; StARD3-GFP cells: median = 1.10, n = 77. **C:** Immunoblotting of FAK and pFAK-Tyr397 in the presence or absence of serum and insulin. Equal amounts of protein were separated by SDS-PAGE and analyzed by immunoblotting with the indicated antibodies. **D:** Quantification of FAK and pFAK-Tyr397 signals from immunoblots. Blots from three independent experiments were quantified by densitometry. Signals were normalized to the signal from control (GFP-overexpressing) cells in full growth medium (+serum, +insulin). Data represent means ± SEM. n = 10 (GFP cells); n = 8 (StARD3-GFP cells). $^{*}P < 0.05$ (Student's t-test), $^{****}P < 0.0001$ (Mann-Whitney-Wilcoxon test) versus GFP control cell line; $^{\dagger}P < 0.05$ for −serum, −insulin versus full serum conditions (Student's t-test). Scale bar = 5 μm.

minutes. Binding of the primary antibody was detected with a PowerVision+ Poly-HRP Histostaining Kit from ImmunoVision Technologies (Norwell, MA), following the manufacturer's recommendations. For both StARD3 and Src, cytoplasmic staining intensity was graded as negative, low, or high when at least 10% of the tumor cells showed protein expression (as is standard in the field). Nuclear staining was not considered relevant for this study. Example images of stained tissue were taken with a Leica DM LB microscope (Leica Microsystems) with an N Plan 20× objective, equipped with an Olympus DP50 color camera (Olympus Corporation).

For ErbB2 (HER-2) staining in the FinProg series, the sections were deparaffinized, followed by antigen retrieval (autoclave treatment at 121°C for 2 minutes in 10 mmol/L sodium citrate buffer, pH 6.0). The primary antibody, CB11 from Novocastra Laboratories (Newcastle upon Tyne, UK), was diluted 1:200 in Powervision blocking solution from Immunovision Inc. (Daly City, CA) and incubated overnight at 4°C. An antimouse-peroxidase polymer from Powervision (30 minutes at room temperature) and diaminobenzidine chromogen were used for visualization. The sections were counterstained with hematoxylin and embedded. Positive and negative control samples (tumors with and without *HER2* amplification in fluorescence *in situ* hybridization) were included in every staining batch. Evaluation of immunohistochemistry was done using a 20× objective. ErbB2 immunoreactivity was scored as negative, low, intermediate, or high, and only high (3+ like) staining was considered as ErB2 positive.[22] In the FinHer series, ErbB2 protein expression was analyzed centrally using the HercepTest, as described elsewhere.[24]

Immunostaining for estrogen receptors was done on adjacent tissue array sections using the monoclonal antibody 6F11 (Novocastra Laboratories; dilution, 1:500) and for progesterone receptors using the antibody 312 (Novocastra Laboratories; dilution, 1:500). Immunostaining was considered as positive when >10% of cancer cells showed staining. The p53 was immunostained with the DO7 antibody (Novocastra Laboratories) at a dilution of 1:500 and Ki-67 using the MM-1 antibody (Novocastra Laboratories; dilution, 1:1000). Ki-67 and p53 staining were classified into three categories: i) negative (no positively staining cancer cell nuclei found), ii) borderline (≤20% nuclei positive), or iii) positive (>20% of cancer cell nuclei stained). Because there was no significant difference in outcome between patients who had either negative or borderline staining, these groups were combined in further analyses.

Quantitative Real-Time RT-PCR

Cells were homogenized in RLT buffer from QIAGEN Sciences (Gaithersburg, MD), and total RNA was isolated with the RNeasy Mini Kit (74104) from QIAGEN Sciences, according to the manufacturer's instructions. The total RNA (1 μg) was reverse transcribed, as previously described.[25] 3-Hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) cDNA was quantified using LightCycler 480 Probe 85 (04689097001) from Roche Diagnostics (Basel, Germany). TATAA binding

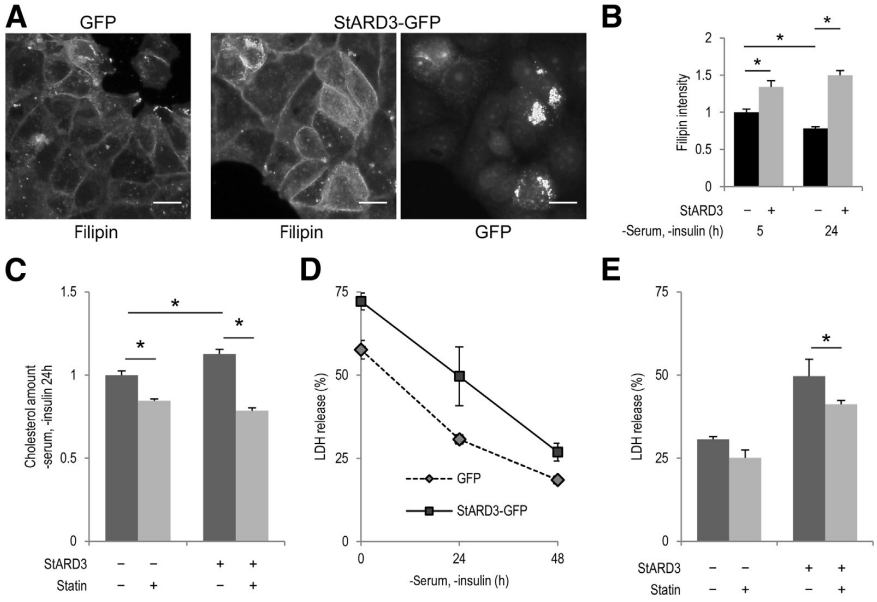


Figure 5 Effects of StAR-related lipid transfer protein 3 (StARD3) overexpression on cholesterol balance in serum- and insulin-deprived conditions. **A:** Filipin staining of cells deprived of serum and insulin for 24 hours. Cells were fixed with paraformaldehyde, stained with filipin, and imaged by wide-field fluorescence microscopy. Please note higher filipin intensity in cells with prominent StARD3—green fluorescent protein (GFP) expression. **B:** Fluorescence intensity quantification from cells deprived of serum and insulin. GFP and StARD3-GFP cells were deprived of serum and insulin for 5 or 24 hours, fixed, stained with filipin, and imaged as in **A**, using the same settings for all conditions. The total filipin fluorescence intensity per cell was measured for 30 to 40 cells per condition and normalized to control cells deprived of serum and insulin for 5 hours. **C:** Total cholesterol content in cells deprived of serum and insulin. GFP and StARD3-GFP cells were deprived of serum and insulin in the presence or absence of lovastatin for 24 hours before lipid analysis. **D:** Sensitivity of cells to streptolysin O (SLO) permeabilization on serum and insulin deprivation for 0, 24, or 48 hours. Resistance of the plasma membrane to SLO was measured as the fraction of total cellular lactate dehydrogenase (LDH) released to the medium after 30 minutes of treatment at 37°C. **E:** Sensitivity of cells to SLO permeabilization on serum and insulin deprivation for 24 hours in the presence or absence of lovastatin. **D:** Values in the presence of lovastatin were normalized to the values in the absence of lovastatin at 24 hours of serum and insulin deprivation. Data represent means ± SEM (**B**); data represent means ± SEM of cholesterol (ng) per protein (μg) (**C**); data represent means ± SD of triplicate measurements (**D**). $n = 6$ to 10 from three independent experiments, cholesterol amounts normalized to GFP control within each experiment (average of controls, 15.51 ng cholesterol/μg protein) (**C**). *$P < 0.05$ (Student's t-test). Scale bar = 20 μm.

box protein was used as an internal control for normalization. Human adrenal tissue was used as a positive control for the cell and breast cancer tissue samples, and RNA and cDNA were generated as for cells. Each sample was run in triplicate. Normalized gene expression values were obtained using LightCycler 480 software version 1.5.1 from Roche Applied Science (Mannheim, Germany), using the basic relative quantification method. The analysis of the patient samples was done identically, but RNA was extracted by using the Versant kPCR Molecular System (Siemens AG, Erlangen, Germany) and normalized gene expression values were obtained using the LightCycler advanced quantification method, as described in detail previously.[26] Adrenal tissue RNA served as a calibrator sample and was used to construct the standard curves for the target and the reference genes.

Statistical Analysis

All data assumed to have a normal distribution were represented as bar charts, with the height of the bar representing the mean and the error bar representing the SEM. Tests of the statistical significance of normally distributed values were performed using a nonpaired two-tailed Student's t-test. Continuous distributions of nonnormally distributed values were summarized in box-and-whiskers plots, with the bottom and top of the box representing the first and third quartiles, respectively, and the band inside the box representing the median. The ends of the whiskers represent the lowest datum still within 1.5 times the interquartile range of the lower quartile, and highest datum still within 1.5 times the interquartile range of the upper quartile. Data points outside of these ranges were not included in the box plots. For nonnormally distributed values, statistical significance was calculated using the Mann-Whitney-Wilcoxon rank sum test. Statistical significance for frequency tables was calculated using the χ^2 test.

An analysis of breast cancer–specific survival was available for the patients included in the FinProg series, and was calculated from the date of breast cancer diagnosis to death from breast cancer, censoring patients who were alive on the date of the last follow-up and those who died from a competing cause on the date of death. In the FinHer series,

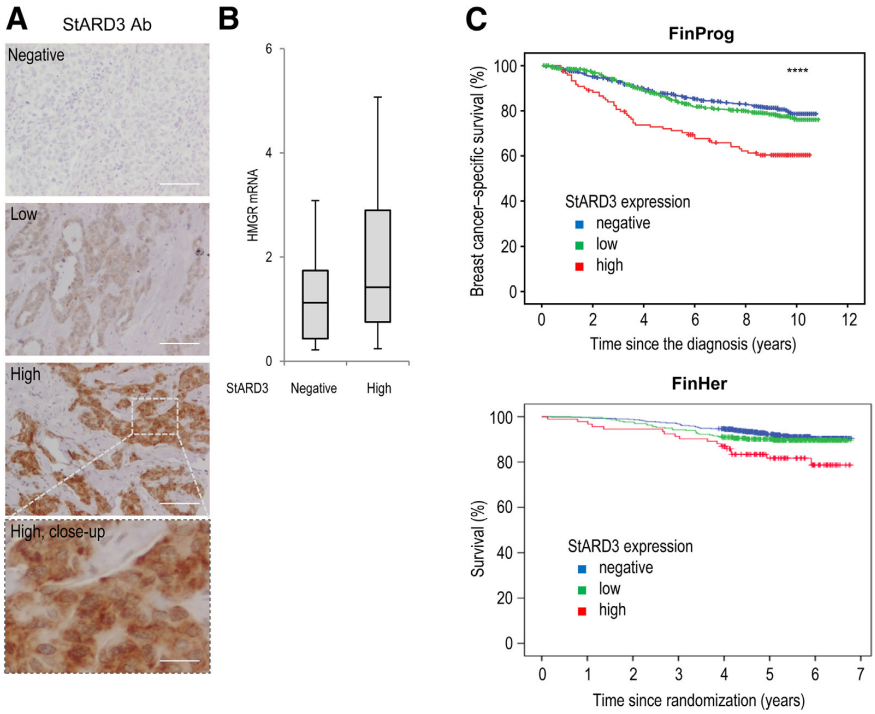


Figure 6 High StAR-related lipid transfer protein 3 (StARD3) protein levels in breast cancer correlate with increased 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) mRNA levels and decreased patient survival. **A:** StARD3 immunohistochemistry of primary breast cancers. **Top three panels:** Exemplary micrographs of negative, low, or high StARD3 levels. **Bottom panel:** a close-up of the indicated area from high expression level. **B:** HMGR mRNA levels in tumors with negative or high StARD3 expression. Total RNA extracted from cells was reverse transcribed, and HMGR cDNA was quantified by real-time RT-PCR. Bars represent HMGR transcript levels and are normalized using HMGR levels in control tissue (adrenal cortex). Box plots: StARD3-negative samples ($n = 31$, median = 1.12) and high StARD3 samples ($n = 38$, median = 1.42). $P = 0.048$ (Mann-Whitney-Wilcoxon test). Outliers, as defined in *Materials and Methods*, are not shown on the plot. **C:** Kaplan-Meier analysis for breast cancer–specific survival of patients in FinProg cohort with negative, low, or high StARD3 protein levels and Kaplan-Meier analysis for end point overall survival of patients with StARD3 protein levels available in FinHer cohort. Negative, low, or high StARD3 expression. $P = 0.007$ (log-rank test) when all three categories are included. ****$P < 0.0001$ (log-rank test) when all three categories are included. Scale bar = 100 μm (**A**); 25 μm (**bottom panel**, **A**).

overall survival was calculated from the date of randomization to the date of death, censoring patients who were alive on the date of the last follow-up visit. Survival between groups was calculated using the log-rank test. The *P* values are two sided.

Results

Stable Overexpression of StARD3 in MCF-7 Cells Alters Cell Morphological Features

To address the role of StARD3 protein in breast cancer cells and tissues, we generated polyclonal rabbit antibodies against the START domain of human StARD3. The affinity-purified antibodies were specific for StARD3 because the immunoreactivity was significantly reduced on transient StARD3 silencing from KPL-4 and BT-474 cells, both of which express StARD3 and ErbB2 at relatively high levels[27] (Supplemental Figure S1, A–D). Furthermore, we found that knockdown of StARD3 in BT-474 cells decreased cell survival (Supplemental Figure S1, E and F). As expected, ErbB2 silencing also decreased cell survival (Supplemental Figure S1, E and F). These observations suggest that both ErbB2 and StARD3 promote the survival of breast cancer cells, and are in agreement with earlier findings.[4]

To gain insights into how StARD3 may promote cell survival independently of *HER2* amplification, we generated MCF-7 (HER2-negative) breast cancer cells that stably overexpress StARD3-GFP or soluble GFP as a control. Overexpression of StARD3-GFP and GFP proteins was verified by Western blot analysis from cell lysates (Figure 1A). Immunoblotting also revealed that the ErbB2 protein levels were not increased on StARD3-GFP overexpression (Figure 1A). Thus, in this cell line, the effects of StARD3 overexpression can be addressed in the absence of HER2 overexpression. Confocal microscopy showed that control GFP expression alone was cytosolic, as expected (Figure 1B), and that the overexpressed StARD3-GFP protein exhibited a punctate distribution, colocalizing, in part, with the late endosomal marker, Lamp-1 (Figure 1B).

Table 1 Summary of StARD3 Immunoreactivity in the Analyzed Patient Cohorts, with Number of Tumors and Distribution of StARD3 Intensity

	StARD3 overexpression (%)		
Study	Negative	Low	High
FinProg (n = 1325)	598 (45.1)	607 (45.8)	120 (9.1)
FinHer (n = 895)	507 (56.6)	296 (33.1)	92 (10.3)

StARD3, StAR-related lipid transfer protein 3.

In addition, some of the StARD3-GFP localized to the plasma membrane, as suggested by the prominent surface fluorescence observed in a fraction of cells (Figure 1, C and D). Total internal reflection fluorescence microscopy revealed a punctate pattern of StARD3-GFP fluorescence in the plasma membrane plane, reflective of plasma membrane domains and/or small vesicles immediately underneath. These results are in accordance with our previous observations on StARD3 subcellular localization.[14]

Remarkably, the overall morphological features of the StARD3-GFP–overexpressing cells were strikingly different from those of control GFP cells. The StARD3-GFP cells grew in clusters of rounded-up cells that did not adhere well to the substratum and lacked growth contact inhibition, as demonstrated by phase-contrast and GFP-fluorescence images of living cells (Figure 1C). Such growth pattern is often characteristic of malignant cells.

StARD3 Overexpression Results in Altered Cellular Cholesterol Balance

To study whether the overexpression of StARD3-GFP affects cellular cholesterol homeostasis, we stained StARD3-GFP– and GFP-expressing cells with filipin, a fluorescent sterol-binding compound that stains cholesterol (and other sterols with a 3-carbon OH-group) in cells.[28] We found that StARD3-GFP cell clusters showed increased filipin intensity, especially in the plasma membrane, as compared to control cells (Figure 2A). This suggests the possibility that the cholesterol content of StARD3-overexpressing cells may be elevated. However, on the basis of biochemical cholesterol determination, the amount of total cellular-free cholesterol in StARD3-GFP cells was not increased, but rather slightly decreased, as compared to control cells (Figure 2B). Therefore, the increased filipin staining intensity in StARD3-overexpressing cells more likely reflects an altered cellular cholesterol distribution. This idea was corroborated by the finding that the level of HMGR mRNA was increased in StARD3-overexpressing cells (Figure 2C). HMGR, the rate-limiting enzyme of cholesterol biosynthesis, is located in the ER and known to be up-regulated on ER cholesterol depletion.[29] To further study if the ER of StARD3-overexpressing cells is cholesterol poor, we analyzed the levels and processing of the key transcriptional regulator of cholesterol metabolism, sterol regulatory element-binding protein 2. Under ER cholesterol depletion, sterol regulatory element-binding protein 2 is increased and its processing to the mature, transcriptionally active form is enhanced. We found this to be the case in StARD3-GFP–overexpressing cells compared to control cells (Figure 2, D and E). Thus, in StARD3-overexpressing cells, the ER appeared to be cholesterol poor and cholesterol synthesis elevated.

To investigate whether the plasma membrane cholesterol content was altered on StARD3 overexpression, we measured the resistance of the cells to SLO. This sterol-binding bacterial toxin forms pores in cholesterol-rich plasma membrane regions.[30] The fraction of cytosolic LDH activity released from cells is used as a measure of membrane permeabilization and, indirectly, as a measure for membrane cholesterol content. As expected, in both GFP- and StARD3-GFP–overexpressing cells, SLO induced the release of LDH from cells (Figure 2F). More important, LDH release from StARD3-GFP–overexpressing cells was higher than that from GFP-expressing cells (1.16-fold; average normalized to GFP control in both SLO concentrations, SD = 0.076). This strongly suggests that the plasma membrane cholesterol content of StARD3-overexpressing cells is higher than that of control cells. Together, these results provide evidence that StARD3 overexpression leads to alterations in cellular cholesterol distribution and homeostatic control, with increased plasma membrane cholesterol but decreased ER cholesterol.

The Effects of StARD3 Overexpression on Cell Morphological Features Are Serum and Insulin Dependent

While studying the morphological effects of StARD3-GFP overexpression, we noticed a striking phenotypic change in StARD3-GFP cells when serum and insulin were deprived from the culture medium. Already within 5 hours after the removal of these growth stimuli, more than half of the StARD3-overexpressing cells became clearly more adherent (Figure 3A). In 24 hours, most of the StARD3-GFP cells had adhered tightly to the substratum and their morphological appearance was more closely reminiscent of that of control GFP cells (Figure 3A). In contrast, serum and insulin withdrawal had no apparent effects on the morphological features of the GFP cells within this time frame (Figure 3A). This change in the morphological features of StARD3-overexpressing cells did not result from reduced StARD3-GFP protein levels on serum and insulin deprivation, as evidenced by Western blot analysis (Figure 3B). In fact, the StARD3-GFP protein levels increased under these conditions (Figure 3, B and C), probably because of increased protein stability, considering that the StARD3-GFP expression was under a viral promoter. Because the proto-oncogene Src is a key regulator of cell adhesion and implicated in breast cancer development,[31] we analyzed its levels. This showed that Src levels were moderately

Table 2 Association of StARD3 Immunoreactivity with Histopathological Parameters in FinProg Material

Parameter	StARD3 overexpression (%)		
	Negative	Low	High
ErbB2 immunostaining			
Negative	503 (93.5)	445 (82.7)	27 (23.1)
Positive	35 (6.5)	93 (17.3)	90 (76.9)
ERBB2-CISH			
Negative	491 (92.8)	437 (82.1)	20 (17.4)
Positive	38 (7.2)	95 (17.9)	95 (82.6)
Molecular subtype			
Luminal A	377 (75.6)	344 (69.4)	14 (13.7)
Luminal B	23 (4.6)	50 (10.1)	34 (33.3)
$HER2^{+}/HR^{-}$	12 (2.4)	40 (8.1)	48 (47.1)
Basal-like	60 (12.0)	35 (7.1)	1 (1.0)
Nonexpressor	27 (5.4)	27 (5.4)	5 (4.9)
pSrc			
Negative	504 (90.8)	422 (76.6)	85 (73.9)
Positive	51 (9.2)	129 (23.4)	30 (26.1)
Estrogen receptor			
Negative (<20%)	155 (29.1)	159 (29.7)	66 (60.0)
Positive (>20%)	377 (70.9)	376 (70.3)	44 (40.0)
p53			
Negative (<20%)	420 (84.5)	383 (80.3)	63 (61.8)
Positive (>20%)	77 (15.5)	94 (19.7)	39 (38.2)
PgR			
Negative (<20%)	226 (41.9)	234 (44.3)	84 (76.4)
Positive (>20%)	313 (58.1)	294 (55.7)	26 (23.6)
Ki-67			
Negative (<20%)	341 (68.2)	309 (60.8)	38 (38.4)
Positive (>20%)	159 (31.8)	199 (39.2)	61 (61.6)
Tumor size distribution*			
Min	1 (0.6)		3 (2.5)
Max	140 (87.0)		90 (76.3)
Median	20 (12.4)		25 (21.2)
Age distribution*			
Min	23 (12.9)		27 (15.9)
Max	96 (53.9)		90 (52.9)
Median	59 (33.1)		53 (31.2)
Nodal metastases ($P = 0.014$)			
No	384 (65.4)	368 (62.5)	61 (51.3)
Yes	203 (34.6)	221 (37.5)	58 (48.7)
Histological features			
Ductal	433 (72.4)	470 (77.4)	102 (85.0)
Lobular	121 (20.2)	69 (11.4)	8 (6.7)
Other	44 (7.4)	68 (11.2)	10 (8.3)
Grade			
I	118 (27.1)	108 (23.5)	6 (6.1)
II	212 (48.7)	211 (46.0)	53 (53.5)
III	105 (24.1)	140 (30.5)	40 (40.4)

$P < 0.0001$, in all parameters except nodal metastases; Mann-Whitney test for statistical significance was used for tumor size and age distributions.

*StARD3 columns negative and low combined.

CISH, chromogenic *in situ* hybridization; HR, hormone receptor; Max, maximum; Min, minimum; PgR, progesterone receptor; StARD3, StAR-related lipid transfer protein 3.

elevated in StARD3-overexpressing cells in complete medium conditions relative to control cells and stayed high in both control and StARD3-overexpressing cells when serum and insulin were removed (Figure 3D).

StARD3-Overexpressing Cells Exhibit Altered FAs and FA Signaling in Serum- and Insulin-Deprived Conditions

To further investigate the molecular basis for the differences in the adherence between GFP and StARD3-GFP cells, we imaged cellular FAs. The StARD3-overexpressing cells were largely rounded up and depleted of FAs in complete medium (Figures 1C and 3A). However, these cells developed prominent, irregularly shaped adhesions under serum and insulin deprivation, as revealed by anti-FAK antibody staining (Figure 4A). Quantification of the area occupied by antibody staining showed that the mean area of FAs was roughly doubled and the variability in FA size was increased on StARD3 overexpression (Figure 4B). Considering that FAK is a key substrate of Src kinase, we analyzed the levels and phosphorylation status of FAK by immunoblotting. This showed that in control cells, there was no substantial difference in FAK protein levels or phosphorylation at Tyr397 on serum and insulin withdrawal (Figure 4, C and D). Although there was a tendency for increased FAK signal, the level and degree of Tyr397 phosphorylation of FAK were not significantly different between control and StARD3-overexpressing cells in complete medium. However, on serum and insulin removal, both the total levels of FAK and its phosphorylation at Tyr397 were significantly increased in StARD3-overexpressing cells compared to control cells (Figure 4, C and D). These results suggest that altered signaling at FAs contributes to the improved adherence of StARD3-overexpressing cells on removal of the growth stimuli.

StARD3-Overexpressing Cells Maintain Increased Cellular and Plasma Membrane Cholesterol Content in Serum- and Insulin-Deprived Conditions

Considering that StARD3 overexpression altered the plasma membrane cholesterol content and that this is known to affect cell adhesion,[32] we investigated how serum and insulin deprivation affected the plasma membrane cholesterol content in control and StARD3-overexpressing cells. The cells were incubated for 5 or 24 hours in the absence of these growth stimuli and stained with filipin, and the fluorescence intensity of individual adherent cells was analyzed. This was considerably easier to perform now that the StARD3 cells were no longer in rounded-up clusters and the cell boundaries were better visualized.

We found that cells overexpressing StARD3-GFP at high levels showed a robust increase in filipin intensity compared to control GFP cells or cells that overexpressed StARD3-GFP at more moderate levels (Figure 5, A and B). This effect was evident both at 5 and 24 hours of serum and insulin deprivation (Figure 5B). Moreover, although the filipin intensity of control GFP cells decreased from 5 to 24 hours, that of StARD3-GFP cells did not (Figure 5B). These results suggest that the plasma membrane cholesterol content of control cells decreased when they were serum deprived, in

Table 3 Association of StARD3 Immunoreactivity with Histopathological Parameters in FinHer Material

Parameter	StARD3 overexpression (%)		
	Negative	Low	High
HercepTest			
0 or 1+	473 (96.7)	168 (58.1)	3 (3.3)
2+	12 (2.5)	73 (25.3)	26 (28.6)
3+	4 (0.8)	48 (16.6)	62 (68.1)
ERBB2-CISH			
Negative	483 (95.3)	199 (67.2)	9 (9.8)
Positive	24 (4.7)	97 (32.8)	83 (90.2)
pSrc			
Negative	361 (64.3)	105 (47.3)	19 (23.2)
Positive	200 (35.7)	117 (52.7)	63 (76.8)
Estrogen receptor			
Negative (<20%)	110 (21.7)	85 (28.7)	58 (63.0)
Positive (>20%)	397 (78.3)	211 (71.3)	34 (37.0)
p53			
Negative (<20%)	379 (79.1)	209 (72.6)	46 (51.1)
Positive (>20%)	100 (20.9)	79 (27.4)	44 (48.9)

$P < 0.0001$ in all cases.

CISH, chromogenic *in situ* hybridization; StARD3, StAR-related lipid transfer protein 3.

accordance with the notion that serum lipoproteins are a major source of cellular membrane cholesterol. In contrast, the StARD3-GFP cells seemed to be capable of maintaining elevated plasma membrane cholesterol levels in the absence of serum and insulin.

Biochemical cholesterol determinations showed that serum- and insulin-starved StARD3-GFP cells maintained an elevated cholesterol content compared to control GFP cells (Figure 5C). To test if this was due to increased cholesterol synthesis, we treated the cells with lovastatin, an inhibitor of HMGR, during the 24-hour serum and insulin deprivation. Statin treatment abolished the capability of StARD3-GFP cells to elevate their cholesterol content (Figure 5C), implying that increased cholesterol synthesis accounted for the elevated cholesterol levels in StARD3-GFP–expressing cells. Moreover, the plasma membrane cholesterol amounts of StARD3-GFP cells were higher than those of control cells after 24 to 48 hours of serum and insulin deprivation, as assessed by sensitivity to SLO permeabilization (Figure 5D). This was also repressed by statin, as expected (Figure 5E).

High StARD3 Protein Levels in Human Breast Tumor Samples Associate with *ERBB2* Amplification, Cholesterol Imbalance, and Src Activation

To investigate StARD3 protein expression in Finnish breast cancer patients, we performed immunohistochemical staining of 1325 primary breast cancer samples deriving from the FinProg cohort[22] using our anti-StARD3 affinity-purified antibody. The StARD3 staining intensity was graded as negative, low, or high (Figure 6A). Stromal cells were not stained, and normal epithelial cells showed negative or low staining. In cells with high staining intensity, the staining exhibited a perinuclear, dot-like staining pattern, with nuclei devoid of staining (Figure 6A). These observations are in concordance with previous results.[8,33] We found that 120 (9.1%) of the samples exhibited high StARD3 staining, 607 (45.8%) showed low StARD3 staining, and 598 (45.1%) were negative (Table 1). High StARD3 immunoreactivity was associated with ErbB2 immunoreactivity and *HER2* amplification, and several factors related to a poor disease outcome, such as high proliferation rate, bigger tumor size, and nodal metastases at the time of diagnosis, as well as estrogen and progesterone receptor negativity and p53 protein expression (Table 2). Considering that our *in vitro* data suggested high StARD3 expression to be associated with increased Src-FAK signaling, we also analyzed the tumor samples by anti-Src Tyr416 (pSrc) staining, which provides a measure of Src activation. We found that in the FinProg material, 1011 (82.8%) of the samples scored as negative and 210 (17.2%) positive for pSrc staining, and that positive Src phosphorylation associated with high StARD3 protein expression (Table 2).

To address StARD3 and pSrc immunoreactivity in an independent patient cohort, we analyzed 895 primary breast cancer tissue specimen from the FinHer study.[23] We found that 92 (10.3%) of these samples exhibited high StARD3 staining, 296 (33.1%) showed low StARD3 staining, and 507 (56.6%) were negative (Table 1). For pSrc, 485 (56.1%) of the FinHer samples were negative, and 380 (43.9%) were positive. In this cohort, high StARD3 immunostaining was strongly associated with *HER2* gene amplification, ErbB2 immunoreactivity, and pSrc positivity, as well as hormone receptor negativity and p53 expression (Table 3). From a subset of the FinHer material, RNA was available and we could analyze HMGR mRNA levels. The subset consisted of 38 primary breast tumors scored as highly StARD3 positive and 31 tumors that were scored StARD3 negative by immunostaining. The analysis showed that the highest HMGR transcript levels were found in the highly StARD3-positive tumors (Figure 6B). Together, these findings suggest that elevated StARD3 protein expression was accompanied by increased cholesterol synthesis and Src activation in human breast cancers. Finally, our analyses showed that high StARD3 protein level was strongly associated with decreased cumulative patient survival in both FinProg ($P < 0.0001$) and FinHer ($P = 0.007$) cohorts (Figure 6C).

Discussion

This study was conducted to address how elevated StARD3 protein expression affects breast cancer cells and to delineate parameters characteristic of primary human breast cancers that exhibit high StARD3 protein content. This is of interest because StARD3 has been shown to be coamplified and co-overexpressed as part of the *HER2/ERBB2* 17q12 amplicon in a subset of breast cancers.[4,17,33–35] Moreover, anti-ErbB2 antibody therapy is widely used in the clinical

setting, yet 50% or even more of ErbB2-positive patients show no response, or become resistant to the treatment.[36] It has been proposed that some of the coamplified gene products may contribute to cancer progression.[3,4,7] StARD3 is an attractive candidate for such a modulator, because as earlier data and our current results in two Finnish patient cohorts indicate, the gene is practically invariably coamplified and the protein co-overexpressed with ErbB2 in *HER2*-positive breast cancers.

StARD3 is a cholesterol-binding protein with a 1:1 stoichiometry[10] and has been shown to affect intracellular membrane and cholesterol transport.[11,14,15] Cancers, including breast cancer, have been associated with alterations in cholesterol metabolism,[37] with both cholesterol and its metabolites being capable of deregulating breast cancer cells.[38] Cancer cells require a continued supply of cholesterol for membrane biogenesis, to support cell growth and division. In addition, cholesterol plays an important role in organizing cholesterol-sphingolipid–rich membrane domains, lipid rafts, that are involved (e.g., in cell signaling and adhesion).[39] Alterations in the plasma membrane cholesterol content affect cell signaling, adhesion, and motility.[40–42] These processes are fundamental to the survival and growth of cancer cells as well as their ability to metastasize.

To dissect the potential ErbB2-independent role of elevated StARD3 protein content in breast cancer cells, we generated MCF-7 cells overexpressing StARD3. On the basis of our results, overexpressing StARD3 caused several alterations in cellular cholesterol homeostasis that are potentially cancer promoting. The StARD3-overexpressing cells had increased levels of the mRNA encoding the rate-limiting cholesterol biosynthesis enzyme HMGR. Increased cholesterol biosynthesis allows cells to survive and to keep dividing, especially in low nutrient conditions, where membrane biogenesis is normally limited. In addition, the StARD3-overexpressing cells exhibited increased plasma membrane cholesterol levels. Moreover, the cholesterol content of StARD3-overexpressing cells was refractory to the withdrawal of the exogenous cholesterol source (ie, serum lipoproteins). Compared to control cells, the StARD3-overexpressing cells managed to maintain elevated total and plasma membrane cholesterol levels in the absence of serum due to the up-regulated cholesterol biosynthesis. The sterol-binding compounds used, filipin and SLO, also bind cholesterol precursor sterols that are up-regulated during cholesterol biogenesis. Thus, these may contribute to the increased sterol content.

A high plasma membrane cholesterol level supports raft-dependent signaling via lipid-anchored proteins, such as Src family kinases. This, in turn, alters the adhesive properties of cells.[43] Indeed, we observed that StARD3 overexpression affected cell adherence: the cells were globular and loosely attached to the substratum in the presence of serum and insulin, but became adhesive on their withdrawal. Consistently, the cells displayed exaggerated FAs and increased phosphorylation of FAK, a major Src substrate. It could be envisioned how such alterations in cell morphological features might, on one hand, favor shedding of cancer cells to the bloodstream and, on the other hand, allow cells to adhere tightly to the underlying tissue and endure limited nutrient conditions within a solid tumor.

There are several mechanisms by which StARD3 could increase plasma membrane cholesterol. First, the newly synthesized StARD3 protein moves to the LEs via the plasma membrane.[16] Increased StARD3 production is likely to increase the amount of StARD3 in the plasma membrane, where it may enhance cholesterol deposition similarly as we have shown for LEs.[14] Second, increased cholesterol in StARD3-containing LEs may become transported to the plasma membrane. The LEs communicate actively with the plasma membrane, and we have recently demonstrated that LE–to–plasma membrane cholesterol transport is an important regulator of cell migration.[42] Finally, newly synthesized cholesterol is efficiently delivered from the ER to the plasma membrane.[44] Thus, increased cholesterol biosynthetic activity observed in StARD3-overexpressing cells probably also facilitates plasma membrane cholesterol enrichment.

In two large nationwide cohorts, FinProg and FinHer, with a total of >2000 breast cancer patients, we observed that *HER2* gene amplification and ErB2 immunoreactivity were tightly associated with high StARD3 protein expression. This subgroup of highly StARD3-positive tumors was composed of approximately 1 of 10 patients in both cohorts. In fact, we could not identify patients who would have high StARD3 immunoreactivity without *HER2* amplification. Hence, we cannot make statements about the specific role of high StARD3 expression alone in the clinical setting. Nevertheless, we found that high StARD3 protein levels correlated with elevated HMGR mRNA levels in the FinHer cohort, in accordance with the *in vitro* cell data. Moreover, high StARD3 protein levels associated with high levels of Src phosphorylation on the activating Tyr416 in both FinProg and FinHer cohorts. This, together with the cell data that show increased Src, elevated phosphorylation of the downstream kinase FAK, and enlarged FAs on StARD3 overexpression in the absence of *HER2* amplification; these findings fit with the idea that high StARD3 protein levels promote Src signaling.

More important, Src activation appears as a key contributor to trastuzumab resistance in *HER2*-positive cancers,[45,46] because patients with active Src were found to be significantly less responsive to trastuzumab than those without Src activation. Although ErbB2 itself is linked to Src activation, other candidates in the *HER2* amplicon have not, to our knowledge, been directly implicated in this process. Our study provides evidence that elevated StARD3 expression is linked to increased Src activity and points to the possibility that StARD3 may contribute to the aggressive behavior of trastuzumab-resistant tumors. Therefore, it would be warranted to investigate if repression of StARD3 activity (e.g., by using small-molecule inhibitors) might be

useful in *HER2*-positive cancers and could be used to combat trastuzumab resistance.

In conclusion, we show that overexpression of *STARD3*, a gene coamplified with *HER2*, results in increased levels of plasma membrane cholesterol and Src, altered adhesiveness, and FAK signaling in *HER2*-negative breast cancer cells *in vitro*. These findings suggest that elevated StARD3 expression has a potential to contribute to breast cancer cell aggressiveness, by increasing membrane cholesterol and thereby enhancing oncogenic signaling. The clinical correlations were consistent because approximately 10% of the analyzed 2220 human breast cancers had high StARD3 protein levels, which associated with high ErbB2 levels, increased Src activity, and poor patient survival. Inhibition of StARD3 function might, therefore, be beneficial in *HER2*-positive cancers.

Acknowledgments

We thank Anna Uro and Pipsa Kaipainen for expert technical assistance.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.ajpath.2014.12.018*.

References

1. Kauraniemi P, Kallioniemi A: Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. Endocr Relat Cancer 2006, 13:39–49
2. Staaf J, Jonsson G, Ringner M, Vallon-Christersson J, Grabau D, Arason A, Gunnarsson H, Agnarsson BA, Malmstrom PO, Johannsson OT, Loman N, Barkardottir RB, Borg A: High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. Breast Cancer Res 2010, 12:R25
3. Kao J, Pollack JR: RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer 2006, 45:761–769
4. Sahlberg KK, Hongisto V, Edgren H, Mäkelä R, Hellström K, Due EU, Moen Vollan HK, Sahlberg N, Wolf M, Børresen-Dale A-L, Perälä M, Kallioniemi O: The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells. Mol Oncol 2013, 7:392–401
5. Hynes NE, Lane HA: ERBB receptors and cancer: the complexity of targeted inhibitors. Nat Rev Cancer 2005, 5:341–354
6. Yarden Y: Biology of HER2 and its importance in breast cancer. Oncology 2001, 61(Suppl 2):1–13
7. Jacot W, Fiche M, Zaman K, Wolfer A, Lamy P-J: The HER2 amplicon in breast cancer: topoisomerase IIA and beyond. Biochim Biophys Acta 2013, 1836:146–157
8. Moog-Lutz C, Tomasetto C, Régnier CH, Wendling C, Lutz Y, Muller D, Chenard MP, Basset P, Rio MC: MLN64 exhibits homology with the steroidogenic acute regulatory protein (STAR) and is overexpressed in human breast carcinomas. Int J Cancer 1997, 71:183–191
9. Alpy F, Tomasetto C: MLN64 and MENTHO, two mediators of endosomal cholesterol transport. Biochem Soc Trans 2006, 34:343–345
10. Tsujishita Y, Hurley JH: Structure and lipid transport mechanism of a StAR-related domain. Nat Struct Biol 2000, 7:408–414
11. van der Kant R, Zondervan I, Janssen L, Neefjes J: Cholesterol-binding molecules MLN64 and ORP1L mark distinct late endosomes with transporters ABCA3 and NPC1. J Lipid Res 2013, 54:2153–2165
12. Alpy F, Latchumanan VK, Kedinger V, Janoshazi A, Thiele C, Wendling C, Rio MC, Tomasetto C: Functional characterization of the MENTAL domain. J Biol Chem 2005, 280:17945–17952
13. Kishida T, Kostetskii I, Zhang Z, Martinez F, Liu P, Walkley SU, Dwyer NK, Blanchette-Mackie EJ, Radice GL, Strauss JF 3rd: Targeted mutation of the MLN64 START domain causes only modest alterations in cellular sterol metabolism. J Biol Chem 2004, 279:19276–19285
14. Hölttä-Vuori M, Alpy F, Tanhuanpää K, Jokitalo E, Mutka A-L, Ikonen E: MLN64 is involved in actin-mediated dynamics of late endocytic organelles. Mol Biol Cell 2005, 16:3873–3886
15. Charman M, Kennedy BE, Osborne N, Karten B: MLN64 mediates egress of cholesterol from endosomes to mitochondria in the absence of functional Niemann-Pick Type C1 protein. J Lipid Res 2010, 51:1023–1034
16. Zhang M, Liu P, Dwyer NK, Christenson LK, Fujimoto T, Martinez F, Comly M, Hanover JA, Blanchette-Mackie EJ, Strauss JF 3rd: MLN64 mediates mobilization of lysosomal cholesterol to steroidogenic mitochondria. J Biol Chem 2002, 277:33300–33310
17. Vinatzer U, Dampier B, Streubel B, Pacher M, Seewald MJ, Stratowa C, Kaserer K, Schreiber M: Expression of HER2 and the coamplified genes GRB7 and MLN64 in human breast cancer: quantitative real-time reverse transcription-PCR as a diagnostic alternative to immunohistochemistry and fluorescence in situ hybridization. Clin Cancer Res 2005, 11:8348–8357
18. Rasband WS: ImageJ. Bethesda, MD, National Institutes of Health, 1997-2012
19. Bligh EG, Dyer WJ: A rapid method of total lipid extraction and purification. Can J Biochem Physiol 1959, 37:911–917
20. Holtta-Vuori M, Vainio S, Kauppi M, Van Eck M, Jokitalo E, Ikonen E: Endosomal actin remodeling by coronin-1A controls lipoprotein uptake and degradation in macrophages. Circ Res 2012, 110:450–455
21. Pimplikar SW, Ikonen E, Simons K: Basolateral protein transport in streptolysin O-permeabilized MDCK cells. J Cell Biol 1994, 125:1025–1035
22. Joensuu H, Isola J, Lundin M, Salminen T, Holli K, Kataja V, Pylkkänen L, Turpeenniemi-Hujanen T, von Smitten K, Lundin J: Amplification of erbB2 and erbB2 expression are superior to estrogen receptor status as risk factors for distant recurrence in pT1N0M0 breast cancer: a nationwide population-based study. Clin Cancer Res 2003, 9:923–930
23. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, Alanko T, Kataja V, Asola R, Utriainen T, Kokko R, Hemminki A, Tarkkanen M, Turpeenniemi-Hujanen T, Jyrkkiö S, Flander M, Helle L, Ingalsuo S, Johansson K, Jääskeläinen A-S, Pajunen M, Rauhala M, Kaleva-Kerola J, Salminen T, Leinonen M, Elomaa I, Isola J; FinHer Study Investigators: Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. N Engl J Med 2006, 354:809–820
24. Joensuu H, Sperinde J, Leinonen M, Huang W, Weidler J, Bono P, Kataja V, Kokko R, Turpeenniemi-Hujanen T, Jyrkkio S, Isola J, Kellokumpu-Lehtinen PL, Paquet A, Lie Y, Bates M: Very high quantitative tumor HER2 content and outcome in early breast cancer. Ann Oncol 2011, 22:2007–2013
25. Yan D, Mäyränpää MI, Wong J, Perttilä J, Lehto M, Jauhiainen M, Kovanen PT, Ehnholm C, Brown AJ, Olkkonen VM: OSBP-related protein 8 (ORP8) suppresses ABCA1 expression and cholesterol efflux from macrophages. J Biol Chem 2008, 283:332–340
26. Sihto H, Kukko H, Koljonen V, Sankila R, Böhling T, Joensuu H: Clinical factors associated with Merkel cell polyomavirus infection in Merkel cell carcinoma. J Natl Cancer Inst 2009, 101:938–945
27. Leivonen S-K, Sahlberg KK, Mäkelä R, Due EU, Kallioniemi O, Børresen-Dale A-L, Perälä M: High-throughput screens identify microRNAs essential for HER2 positive breast cancer cell growth. Mol Oncol 2014, 8:93–104

28. Linder MD, Uronen RL, Holtta-Vuori M, van der Sluijs P, Peranen J, Ikonen E: Rab8-dependent recycling promotes endosomal cholesterol removal in normal and sphingolipidosis cells. Mol Biol Cell 2007, 18: 47−56
29. Ikonen E: Cellular cholesterol trafficking and compartmentalization. Nat Rev Mol Cell Biol 2008, 9:125−138
30. Bhakdi S, Tranum-Jensen J, Sziegoleit A: Mechanism of membrane damage by streptolysin-O. Infect Immun 1985, 47:52−60
31. Finn RS: Targeting Src in breast cancer. Ann Oncol 2008, 19: 1379−1386
32. Staubach S, Hanisch F-G: Lipid rafts: signaling and sorting platforms of cells and their roles in cancer. Expert Rev Proteomics 2011, 8: 263−277
33. Cai W, Ye L, Sun J, Mansel RE, Jiang WG: Expression of MLN64 influences cellular matrix adhesion of breast cancer cells, the role for focal adhesion kinase. Int J Mol Med 2010, 25:573−580
34. Alpy F, Boulay A, Moog-Lutz C, Andarawewa KL, Degot S, Stoll I, Rio M-C, Tomasetto C: Metastatic lymph node 64 (MLN64), a gene overexpressed in breast cancers, is regulated by Sp/KLF transcription factors. Oncogene 2003, 22:3770−3780
35. Kauraniemi P, Bärlund M, Monni O, Kallioniemi A: New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res 2001, 61:8235−8240
36. Mohd Sharial MSN, Crown J, Hennessy BT: Overcoming resistance and restoring sensitivity to HER2-targeted therapies in breast cancer. Ann Oncol 2012, 23:3007−3016
37. Danilo C, Frank PG: Cholesterol and breast cancer development. Curr Opin Pharmacol 2012, 12:677−682
38. Silvente-Poirot S, Poirot M: Cancer: cholesterol and cancer, in the balance. Science 2014, 343:1445−1446
39. Simons K, Ikonen E: Functional rafts in cell membranes. Nature 1997, 387:569−572
40. Ramprasad OG, Srinivas G, Rao KS, Joshi P, Thiery JP, Dufour S, Pande G: Changes in cholesterol levels in the plasma membrane modulate cell signaling and regulate cell adhesion and migration on fibronectin. Cell Motil Cytoskeleton 2007, 64:199−216
41. Raghu H, Sodadasu PK, Malla RR, Gondi CS, Estes N, Rao JS: Localization of uPAR and MMP-9 in lipid rafts is critical for migration, invasion and angiogenesis in human breast cancer cells. BMC Cancer 2010, 10:647
42. Kanerva K, Uronen R-L, Blom T, Li S, Bittman R, Lappalainen P, Peränen J, Raposo G, Ikonen E: LDL cholesterol recycles to the plasma membrane via a Rab8a-Myosin5b-actin-dependent membrane transport route. Dev Cell 2013, 27:249−262
43. Patra SK: Dissecting lipid raft facilitated cell signaling pathways in cancer. Biochim Biophys Acta 2008, 1785:182−206
44. Heino S, Lusa S, Somerharju P, Ehnholm C, Olkkonen VM, Ikonen E: Dissecting the role of the golgi complex and lipid rafts in biosynthetic transport of cholesterol to the cell surface. Proc Natl Acad Sci U S A 2000, 97:8375−8380
45. Zhang S, Huang W-C, Li P, Guo H, Poh S-B, Brady SW, Xiong Y, Tseng L-M, Li S-H, Ding Z, Sahin AA, Esteva FJ, Hortobagyi GN, Yu D: Combating trastuzumab resistance by targeting SRC, a common node downstream of multiple resistance pathways. Nat Med 2011, 17:461−469
46. Vu T, Claret FX: Trastuzumab: updated mechanisms of action and resistance in breast cancer. Front Oncol 2012, 2:62

PLOS ONE

RESEARCH ARTICLE

Language-Agnostic Reproducible Data Analysis Using Literate Programming

Boris Vassilev[1]*, Riku Louhimo[2], Elina Ikonen[1,3], Sampsa Hautaniemi[2]

1 Department of Anatomy, Faculty of Medicine, University of Helsinki, Helsinki, Finland, **2** Research Programs Unit, Genome-Scale Biology, University of Helsinki, Helsinki, Finland, **3** Minerva Foundation Institute for Medical Research, Helsinki, Finland

* boris.vassilev@helsinki.fi

Citation: Vassilev B, Louhimo R, Ikonen E, Hautaniemi S (2016) Language-Agnostic Reproducible Data Analysis Using Literate Programming. PLoS ONE 11(10): e0164023. doi:10.1371/journal.pone.0164023

Editor: Frederique Lisacek, Swiss Institute of Bioinformatics, SWITZERLAND

Received: May 27, 2016

Accepted: September 19, 2016

Published: October 6, 2016

Data Availability Statement: The used software package is available from https://github.com/borisvassilev/lir. The presented use case is available from https://github.com/borisvassilev/endobrca.

Funding: BV received personal grants by the Finnish Society of Science and Letters (http://www.scientiarum.fi/eng/); the Biomedicum Helsinki Foundation (http://www.biomedicum.com/index.php?page=112&lang=2); The Paulon Säätiö (http://www.paulo.fi/); the K. Albin Johanssons Stiftelse (http://www.foundationweb.net/johansson/); and the Ida Montinin Säätiö (www.idamontininsaatio.fi/

Abstract

A modern biomedical research project can easily contain hundreds of analysis steps and lack of reproducibility of the analyses has been recognized as a severe issue. While thorough documentation enables reproducibility, the number of analysis programs used can be so large that in reality reproducibility cannot be easily achieved. Literate programming is an approach to present computer programs to human readers. The code is rearranged to follow the logic of the program, and to explain that logic in a natural language. The code executed by the computer is extracted from the literate source code. As such, literate programming is an ideal formalism for systematizing analysis steps in biomedical research. We have developed the reproducible computing tool Lir (**li**terate, **r**eproducible computing) that allows a tool-agnostic approach to biomedical data analysis. We demonstrate the utility of Lir by applying it to a case study. Our aim was to investigate the role of endosomal trafficking regulators to the progression of breast cancer. In this analysis, a variety of tools were combined to interpret the available data: a relational database, standard command-line tools, and a statistical computing environment. The analysis revealed that the lipid transport related genes *LAPTM4B* and *NDRG1* are coamplified in breast cancer patients, and identified genes potentially cooperating with *LAPTM4B* in breast cancer progression. Our case study demonstrates that with Lir, an array of tools can be combined in the same data analysis to improve efficiency, reproducibility, and ease of understanding. Lir is an open-source software available at github.com/borisvassilev/lir.

Introduction

The results of a study can be reproduced and evaluated when all data has been disclosed [1] and the computational methods have been shared in detail [2]. A study of 18 published data analyses showed that the majority of the analyses could not be reproduced, often due to the incomplete specification of the data processing and the analysis [3]. To improve reproducibility of computational analyses several guidelines have been suggested. For example, Sandve et al. proposed a list of ten simple rules for reproducible computational research [4]. Wilson et al.

). RL received no specific funding for this work. EI received grant 282192 by the Academy of Finland (www.aka.fi). SH received funding by Biocentrum Helsinki (http://www.helsinki.fi/biocentrum/). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

compiled an itemized list of best practices for scientific computing [5], and Shade et al. presented a step-by-step guide to computational analysis aimed at biologists [6]. While it would be beneficial to a data analyst to follow stringently all provided guidelines, there is a paucity of computer software that facilitates the implementation of all of them.

Generic software used in programming, such as version control and build utilities, cover some of the needs for reproducible analysis. Other software is specifically aimed at computational data analysis. One such software, Sweave, allows embedding R code into a document typeset with LaTeX [7]. The results of the automated data analysis described in the embedded code are inserted into the generated report to guarantee reproducibility. The utility of Sweave inspired an improvement, Knitr, that addresses most of the perceived shortcomings of its predecessor [8]. Both tools offer an electronic, automated version of the "lab notebook" as described by Noble [9] for the R Statistical Environment [10]. IPython is a notebook solution for the Python programming language [11]. It has evolved into Jupyter (jupyter.org), a platform which supports reproducible computing notebooks in many programming languages, and has become widely accepted [12]. A curated list of publications that employ such notebooks, with links to the data analyses, is available at go.nature.com/mqonbm.

Complex frameworks for the integration of heterogeneous, large-scale biological data have also been developed [13, 14]. An interesting solution proposed by Kitchin [15] addresses the problem of sharing the data analysis in journal publications by embedding the computer executable code within the published PDF.

Existing solutions either assume the exclusive use of a single programming language, such as R or Python [7, 8, 11], or require a non-trivial tool chain and a domain specific language [13, 14]. Here, we introduce Lir: a tool for reproducible computing that encourages and simplifies the use of any combination of existing software platforms and programming languages within the same data analysis [16].

Lir is based on the idea of literate programming as proposed by Donald Knuth [17]. Literate programming allows the user to organize and document their work free of the restrictions on code placement, file structure, or naming imposed by programming languages. This is achieved by embedding all code within named paragraphs called *code chunks*. Code chunks may appear in any order within the literate source file. In the original implementation by Knuth, a code chunk would immediately follow the text that explains the rationale of the code and its importance within the logic of the whole program; Lir follows that convention. Code chunks have descriptive names in natural language, and can contain other code chunks through name references. Thus, all computer code can be presented as the logic of the analysis requires. Additionally, Lir provides facilities for combining programs written in any programming language in the same data analysis. For instance, a relational database can be used to organize and query the data, a scientific computing platform to analyze it, and a statistical computing environment to summarize and visualize it.

The complete analysis—documentation, source code, placeholders for display items, discussion of results—is maintained as a text file. This file (S1 File) contains code chunks in the syntax recognized by the most widely used language-agnostic literate programming tool, `noweb`. This source file is used to dynamically generate all results and the final human-readable document (S2 File).

To demonstrate the application of Lir, we studied the effects of endosomal trafficking regulators on the progression of breast cancer by using gene amplification, mRNA expression, and patient survival data from The Cancer Genome Atlas repository [18]. Endosomal trafficking is the process by which cells internalize, sort, and recycle nutrients and signaling molecules with the help of vesicles formed at the cell's outer membrane, the plasma membrane. Defects in endosomal uptake, sorting, recycling, and degradation of cell metabolites, external signals, and

cell surface receptors can lead to deregulation of the cell cycle. The focus of the study was on genes and gene products involved in endosomal lipid trafficking and in breast cancer progression. The complete analysis is available as a supplement (S1 File).

Methods

Lir

Lir defines a markup language for defining and documenting a data analysis, and presenting its results. The data analysis is defined by the declared data objects (data files), the defined data transformations (executable programs that consume and produce data files), and the declared rules for applying a data transformation on data objects. Lir provides three command-line programs, `lir-tangle`, `lir-make`, and `lir-weave`, that generate the results of the data analysis and compile a human-readable document (Fig 1A). For most use cases, the three steps can be invoked in simple succession; the convenience program `lir` does that. The Tutorial that accompanies the Lir implementation gives an introduction to how Lir is meant to be used, and the User Guide provides a complete reference (available at https://github.com/borisvassilev/lir and https://github.com/borisvassilev/lir-tutorial).

Source file. A Lir source file is a valid `noweb` source file. The `noweb` syntax splits a text file into a sequence of documentation chunks and code chunks, and imposes no restrictions on the content of these chunks [19]. Code chunks are named, and names can be arbitrary strings, containing for example formatted text or mathematical symbols and formulas. Code chunks can include other code chunks through name references. Code chunks that are not referenced by other code chunks are *root* code chunks. The executable programs evaluated during the data analysis are defined in root code chunks named after the transformation object they represent. Documentation chunks can be formatted and structured using any markdown recognized by Pandoc.

Analysis definition. Lir extends the semantics of `noweb` by embedding in code chunks with special names the dependencies between executable code, input data, and results. The data objects, data transformations, and the rules describe a directed acyclic graph (DAG): the data objects are the vertices of the graph, and the rules for applying data transformations are the directed edges. In this DAG, the sources are the input data files; the final results and figures are the sinks. The input data files (the sources) must be explicitly declared. If a placeholder for a display item (for example, a table or a figure) appears in the Lir source file, the file containing the display item is a sink. Additional sinks can be declared for results that are not displayed in the final document but have to be generated. The rules for applying a data transformation on data objects are declared in their own code chunks, in the language recognized by Make. In these rules, data objects are referred to by the names used in the declarations of sources and sinks. Data transformations are referred to by the root code chunk names in which they are defined.

Generating results. In a first step, `lir-tangle` extracts the data transformations defined in the source file to executable programs, and uses the declared data objects and rules to build a valid makefile. In the next step, `lir-make` invokes Make to generate all results, observing the dependencies between input data, intermediate files, and final results. By keeping this step separate it is possible to execute it on a different machine, for example a remote application server. Only the tools used in the data analysis and Make have to be installed on the machine running the data analysis: neither Lir, nor any of its dependencies (`noweb`, SWI-Prolog, Pandoc, Bash, and so on) are required.

Compiling the final document. In the last step, `lir-weave` produces an HTML document that contains the full text of the source file, all display items, and links to all generated

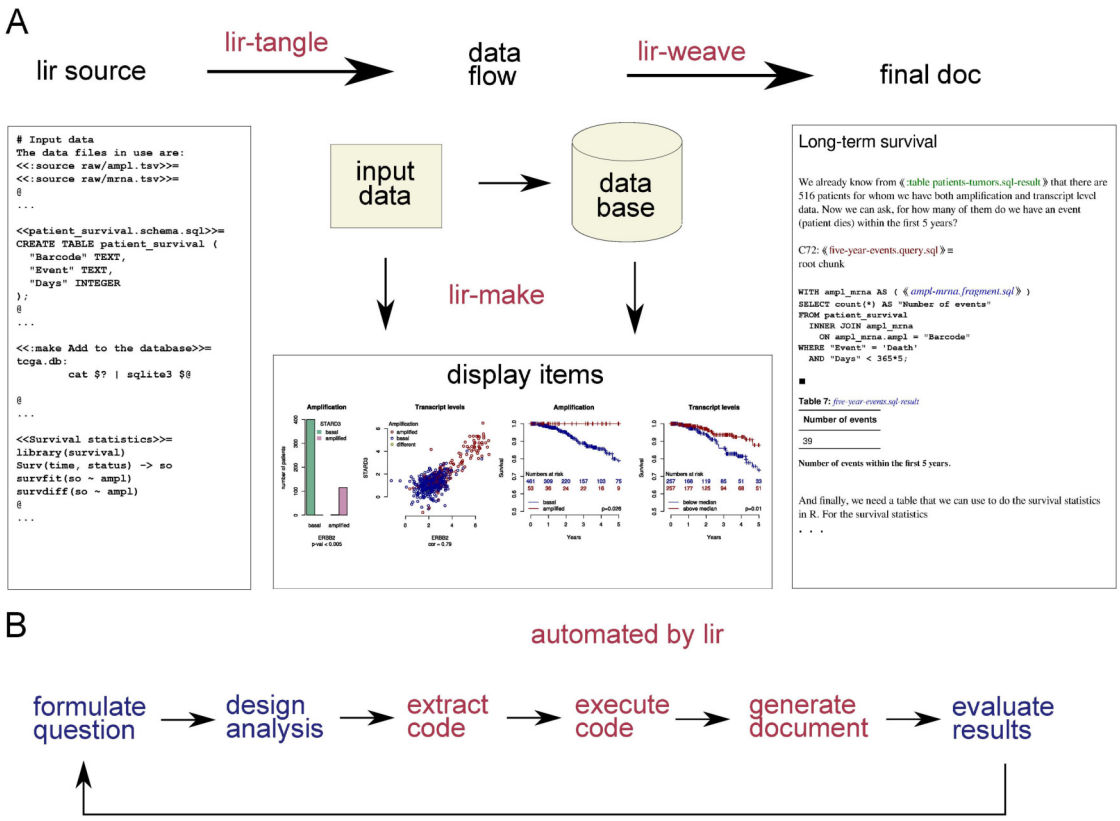


Fig 1. Overview of the method and tools provided by Lir. A: Documented executable code, placeholders for the display items, and discussion of the displayed results are maintained as a text file: the Lir source. Dependencies between the input data (data objects) and the executable code (data transformations) are an integral part of the source. The executable code and instructions for running it are extracted from the source file by `lir-tangle`. The code is evaluated by `lir-make` to obtain results and display items. A human-readable, cross-linked document that faithfully represents the literate source text file and integrates all display items is compiled by `lir-weave`. **B**: An overview of the work flow using this method. All steps that are performed by a computer are automated, including resolving dependencies between data objects and data transformations. This facilitates exploratory data analysis with short iteration cycles between formulating a question, evaluating its answer, and refining or extending the analysis.

doi:10.1371/journal.pone.0164023.g001

files that are not displayed (see S2 File). The code chunks containing the executable programs are numbered and cross-linked to help browsing and reading the code. If markdown is used in the documentation chunks, the final document is formatted accordingly and has a table of contents. Figures, tables, and plain text results are displayed with their own labels ("Figure", "Table", "Listing"), may have captions, and are numbered separately.

Implementation. Tangling the source file with `lir-tangle` is implemented as a Bash script that makes use of the low-level tools provided by `noweb`, and standard Linux command line tools like Awk, `sed`, and `grep`. Generating all results with `lir-make` is implemented as an invocation of Make with a makefile generated by `lir-tangle` and the appropriate command-line options. The implementation of `lir-weave` employs two steps. First, the `lir` source file is converted to an intermediate representation. This is implemented in SWI-Prolog

[20], as it is a convenient tool for both parsing the Lir source file and generating HTML markup. In this step, code chunks and display items are numbered, cross-linked, and structured using HTML `span` and `div` elements modified with HTML attributes. The intermediate representation is compiled to a final document using Pandoc and formatted with a default CSS style sheet provided by Lir. Compiling the final document with `lir-weave` may be extended to produce other output formats, for example a PDF file for printing. We chose HTML as the first supported final document format for two reasons. First, it is a format that can be viewed any platform with a modern web browser. Second, the final layout and formatting can be customized by the user by providing another CSS style sheet, without changing or extending the implementation.

Compatibility. The tools provided by Lir and all its dependencies are open source software that can be installed on any GNU/Linux, BSD, and OS X. It is possible to install and use Lir on a Microsoft Windows system, but the differences in file path naming conventions between POSIX and Microsoft Windows might cause incompatibilities at the level of the Lir source file. We cannot currently guarantee, for example, that a Lir source file that contains file paths with backslashes as component separators or file names with spaces (allowed by the Microsoft Windows Uniform Naming Convention) can be interpreted on a POSIX system without normalizing the file paths.

Breast cancer data

The raw data on gene amplification and mRNA levels for breast cancer patients were obtained from The Cancer Genome Atlas project (TCGA) [18] and pre-processed to obtain two data matrices: the amplification status and mRNA levels for approximately 18 thousand genes for tissue samples from over 500 patients.

We downloaded level 1 Affymetrix SNP Array data from the Cancer Genome Atlas [18]. The data was processed anonymously. All TCGA data were preprocessed using Anduril [13]. We genotyped the probes, and estimated and normalized copy-number values to 2 with the CRLMM algorithm [21]. Copy-number data were segmented with the circular binary segmentation (CBS) algorithm using the R package DNAcopy (parameters `undo.splits = sdundo`, `SD = 3`, `alpha = 0.01`) [22]. Copy-numbers for each gene were assigned to three states (gained, normal and deleted) similarly to TCGA [18].

Gene expression had been measured with an Agilent two-channel microarray from which only the channel containing measurements from a patient sample were used. We mapped probes to genes, and removed probes mapping to multiple genes or no genes. Gene expression values for genes, which are tagged by several probes, were combined using the median over these probes. Gene expression values were normalized to a mean of 0 over the samples.

For each gene and sample, the amplification value was either 0 (no amplification) or 1 (gene is amplified). The mRNA level was represented as a numerical value that can be compared between patient samples for the same gene. Each gene was uniquely identified by an Ensembl Gene ID. Each sample was uniquely identified by a TCGA "barcode" that contains meta-information including, among others, details about the collection site, sample type, and study participant. Clinical data for all patients, including survival data, was also obtained from TCGA.

Statistical methods and visualizations

The overall goal of the case study was to see how well the genes that regulate endosomal trafficking correlate at mRNA, miRNA and copy-number levels. In the copy-number correlation analyses we used the χ^2 test statistic for a 2×2 contingency table that represents the amplification status values for each pair of genes. The contingency table itself was visualized as a bar plot where

each bar represents one of the table cells. The correlation of the mRNA levels was estimated using the Pearson's correlation coefficient. All plots were generated using R's built-in methods except for survival curves, which were generated using R's `survival` package [23, 24].

Results and Discussion

In order to demonstrate Lir, we analyzed breast cancer data including amplification status, mRNA level data, and survival data from 516 patients. The analysis employed a relational database, an array of standard command-line tools for text and table manipulation, and the R statistical environment for statistical analysis and visualization. The results were obtained using the basic work flow outlined in Fig 1B.

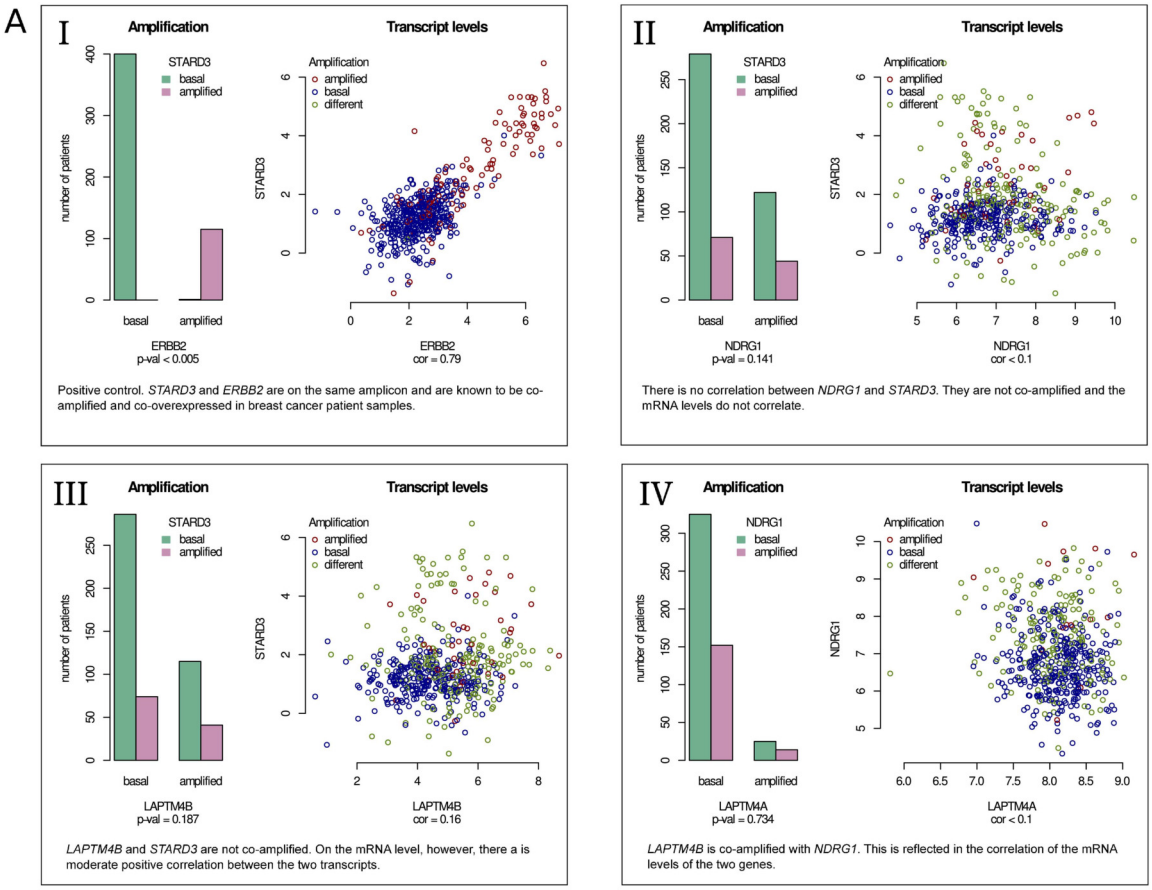
We were initially interested in three genes. These genes code for proteins implicated both in endosomal lipid trafficking and in breast cancer. StARD3 (**StAR**-related lipid transfer **d**omain protein **3**) is a cholesterol-binding late endosomal (LE) protein that contributes to the progression of ErbB2-positive breast cancer, an established aggressive cancer sub-type [25]. NDRG1 (**N**-myc **d**ownstream **r**egulated **g**ene **1**), implicated in several cancers [26], regulates endosomal trafficking and degradation of the cell-surface receptor for low-density lipoprotein [27]. LAPTM4B (**L**ysosomal-**a**ssociated **t**rans**m**embrane **p**rotein **4**-**b**eta) is a LE membrane protein associated with chemotherapy resistance in several cancers. Recent results show that it controls ceramide export from the LE and thereby affects sensitivity to anti-cancer drugs [28]. Altered *STARD3*, *NDRG1*, or *LAPTM4B* expression has been associated with neoplasms in multiple reports. According to experimental findings in cell culture models, all three proteins function at different points along the endosomal trafficking route, and regulate intracellular lipid trafficking. It was therefore of interest to assess if their gene amplification or mRNA expression levels correlate with each other or with breast cancer patient survival.

Correlating genes of interest

The genes of interest, *STARD3*, *NDRG1*, and *LAPTM4B* are located in the following chromosomal regions: *STARD3* in the *ERBB2* amplicon in 17q12-21 [29], *NDRG1* in 8q12-24, often coamplified with MYC [30], and *LAPTM4B* on the same chromosome arm in 8q22 [31], approximately 35 Mb apart from the *NDRG1* locus. First, we investigated whether the gene amplification or mRNA levels of *STARD3*, *NDRG1*, and *LAPTM4B* correlate with each other in breast cancer tumor tissue. As a positive control we calculated correlation between *STARD3* and *ERBB2*, which are known to be highly correlated on the DNA, mRNA, and protein level [25, 29, 32]. The amplification statuses of these two genes is almost identical, and they correlate strongly at the mRNA level (Pearson's $\rho = 0.79$, Fig 2A, Panel I).

We found no correlation between *NDRG1* and *STARD3* amplification status or mRNA levels (Pearson's $\rho < 0.1$, Fig 2A, Panel II). Instead, a weak correlation between *STARD3* and *LAPTM4B* was observed. Although the amplification status of these two genes did not correlate, on the mRNA level there was a weak positive correlation (Pearson's $\rho = 0.16$, Fig 2A, Panel III). This suggests that the two genes might be co-regulated on the transcriptional or post-transcriptional level. Interestingly, *LAPTM4B* and *NDRG1* correlate positively both in their gene amplification status and mRNA levels (Pearson's $\rho = 0.34$, Fig 2A, Panel IV). This may be related to their close proximity in the 8q region amplified in cancers.

An overview of the data flow used to generate the above results is outlined in Fig 2B. A range of tools are used: standard command line tools for examining and preparing the input data (`wc`, `sed`, `awk`, `tr`, `cut`, etc.), a relational data base (SQLite) for the meta data, R for the statistical analysis and plot generation. Code chunks are given descriptive names in free text, allowing for a self-documenting, consistent, literate writing style independent of the



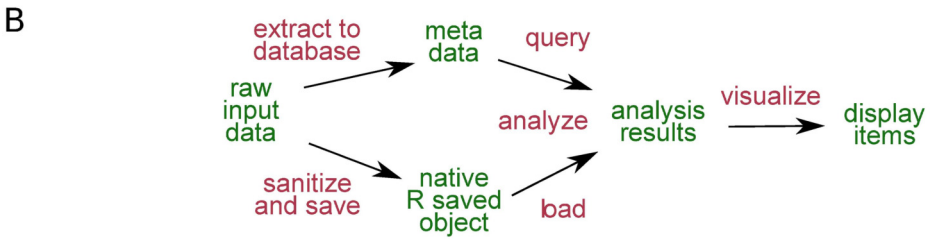


Fig 2. Using Lir to correlate genes of interest. A: Four of the display items generated by the analysis outlined in (A). In the scatter plots, red indicates a sample in which both genes were amplified, blue indicates a sample in which both genes are at basal levels, and green indicates a sample with differential amplification status of the two genes. The interpretation of the results of each display item shown here is added to the source file: the interpretation becomes an integral part of the analysis. **B**: Diagram demonstrating the data flow for an analysis. The input data (a large data matrix in a text file) is sanitized and saved as a native R object. Meta-data of the genes and the samples is saved to a relational database to facilitate querying the data. The relevant data is extracted, analyzed, and visualized, producing several display items. In this diagram, data objects are colored in green, data transformations are colored in red, and arrows represent the dependencies declared in the source file and used by Lir to generate the intermediate data objects and the display items.

doi:10.1371/journal.pone.0164023.g002

programming language used (for example, see S1 File, lines 618–702). Scripting tools like R offer little flexibility when it comes to positioning code within the file or sharing code between files. Multiple references by name to the same code chunk can be used to avoid code repetition and reorganize code, overcoming this practical limitation (see S1 File, line 264, used at 225, 535, 598, 894, . . .).

Analyzing the effect of genes of interest on patient survival

We next assessed whether the overexpression of a gene or a pair of correlated genes have an effect on patient survival within five years. We have earlier reported that high StARD3 protein levels associate with poor breast cancer specific survival in two Finnish nationwide patient cohorts [25]. This was, however, not observed on the amplification or transcript level in the TCGA data set (S2 File). Thus, we tested the survival effect of *SERPINA1*, another recently reported predictor of survival in breast cancer [33].

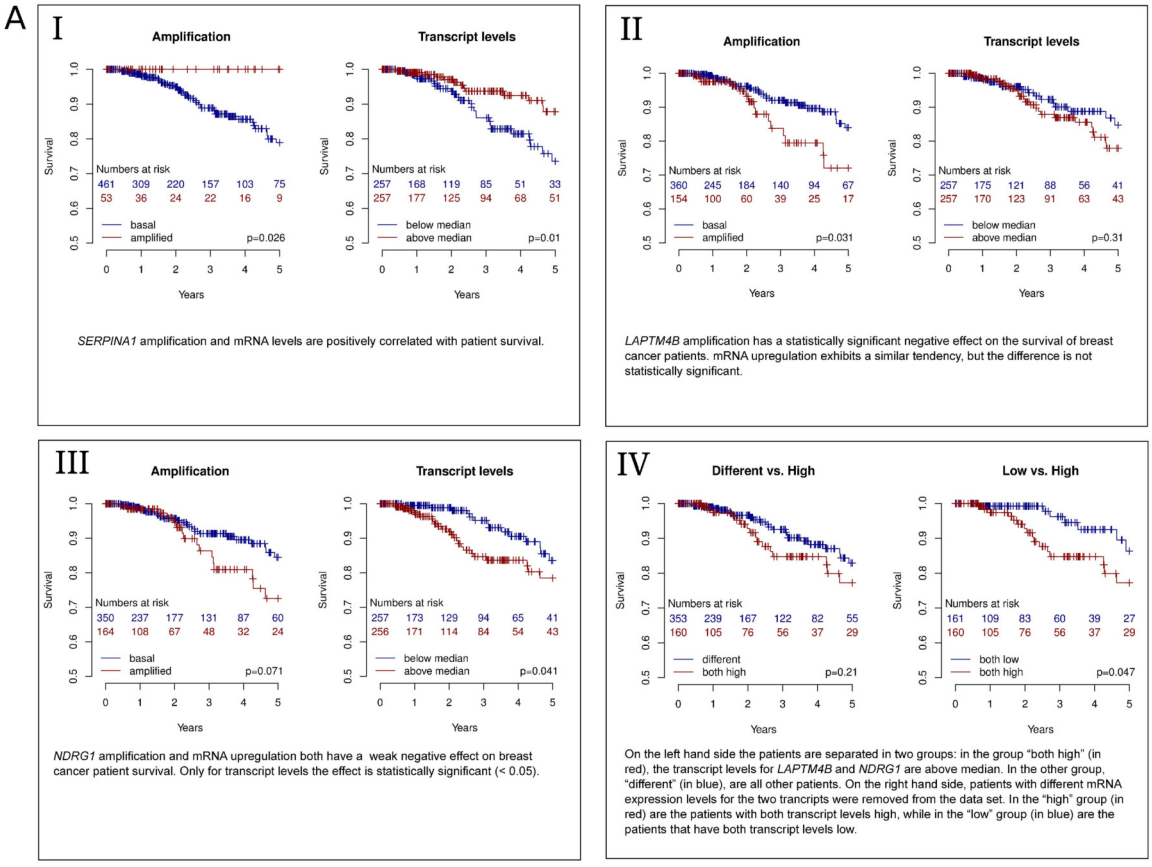
A clear positive correlation of both *SERPINA1* amplification status and mRNA levels with patient survival was observed (Fig 3A, Panel I). On the other hand, both *LAPTM4B* and *NDRG1* levels correlated with reduced patient survival: in the case of *LAPTM4B*, the negative effect was evident at the level of gene amplification (Fig 3A, Panel II), while for *NDRG1*, the correlation was better at the transcript level (Fig 3A, Panel III).

To test the combined effect of *LAPTM4B* and *NDRG1* mRNA overexpression on patient survival, the patients were split into two groups. First, one group contained those patients that have high mRNA levels (over median within the gene) for both genes, and the other group all other patients (Fig 3A, Panel IV). These groups had similar survival. Then, the mixed mRNA level patients (one mRNA below, the other above median) were excluded from the data set and only those patients with high or low mRNA levels for both transcripts were compared with each other. This revealed a lower survival for patients with elevated transcript levels; however, the effect was similar to that obtained by *NDRG1* upregulation alone, implying that co-overexpression of *LAPTM4B* and *NDRG1* did not have an additional negative effect on patient survival in this data set.

The data flow of these analyses is summarized in Fig 3B. The data flow documents how the reuse of existing data objects avoids the need for excess code and lowers the amount of computational work. The structure of the data flow strongly resembles that of the work flow presented in Fig 2B. The formal syntax used by Lir for declaring dependencies allows to reuse existing data flow structures, further avoiding unnecessary work (see S1 File, lines 798–801, reusing lines 285–90, 293–7, and 301–5).

Assessing miRNA target genes by integrating two independent data sets

Micro RNA (miRNA) are short non-coding RNA molecules that regulate mRNAs after transcription, usually inducing gene silencing [34, 35]. Individual miRNAs may target as many as 100 different mRNA molecules. For three of the genes of interest, we identified experimentally validated miRNAs relevant in the context of breast cancer: for *ERBB2*, miR-155 [36], for *LAPTM4B*, miR-188 [37], and for *NDRG1*, miR-769 [38]. To find potential additional targets for each miRNA, we used miRWalk [39], additionally querying four more online resources: microRNA.org [40], miRDB [41], RNA22 [42], and TargetScanHuman [43]. There were 2553 genes in the TCGA breast cancer data set that were predicted targets with the selected cut-off (Fig 4A, Panel I). Of all the predicted targets that correlated with the corresponding gene on the transcript level (S2 File and Fig 4A, Panel II), two of the *LAPTM4B*/miR-188 genes were especially interesting: *PVR* (Poliovirus receptor protein),



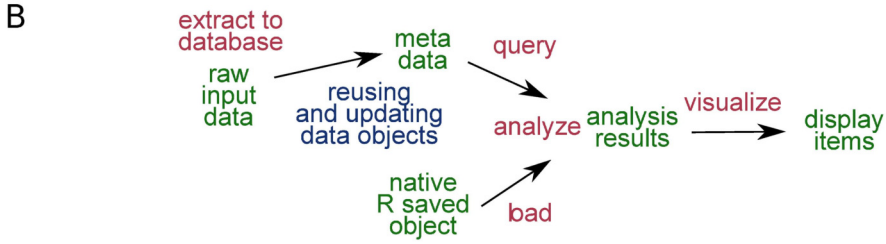


Fig 3. Using Lir to determine the effects of upregulation of the genes of interest on patient survival. A: Three of the display items generated by the analysis outlined in (A). As in Fig 2, the interpretations of the results shown below each display item appear verbatim in the literate source. **B**: Diagram demonstrating the data flow for further analysis based on the results shown in Fig 2. Importantly, data objects generated in the previous analysis are reused. In this diagram, data objects are colored in green, data transformations are colored in red, and arrows represent the dependencies declared in the literate source file and used by Lir to generate the intermediate data objects and the display items.

doi:10.1371/journal.pone.0164023.g003

A

I

```
957 hsa-miR-155-5p
826 hsa-miR-188-5p
770 hsa-miR-769-3p
2553 total
```

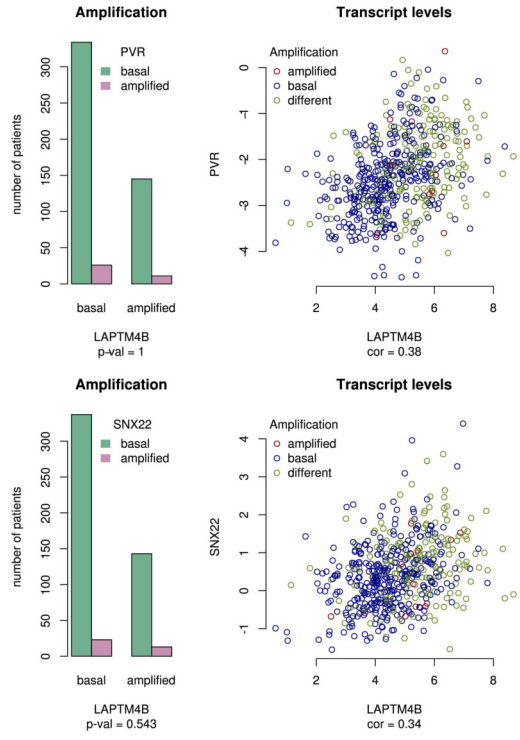
Five different tools for predicting micro-RNA targets were applied to three experimentally validated mirco-RNA sequences. These are the numbers of predicted targets with a strength of at least 4 out of 5, or strength 3 for at least two different sites.

II

```
Gene:  LAPTM4B
miRNA:  hsa-miR-188-5p
Targets:
 [1] "LAPTM4B"   "YWHAZ"     "JMJD6"
 [4] "SLMO2"     "TXNRD1"    "C1orf106"
 [7] "RBL1"      "CDC25B"    "PVR"
[10] "TMEM194A" "PCMT1"     "UQCRB"
[13] "UCK2"      "FBXO45"    "OSGIN2"
[16] "USP31"     "SNX22"     "AQP9"
[19] "SPAST"
```

Using Pearson's correlation as a measure, the association of predicted targets of hsa-miR-188-5p to *LAPTM4B* on the mRNA level was estimated. The table shows all genes that showed a correlation coefficient of > 0.33.

III



Scatter plots of the correlation of *LAPTM4B* to *PVR* and *SNX22*.

B

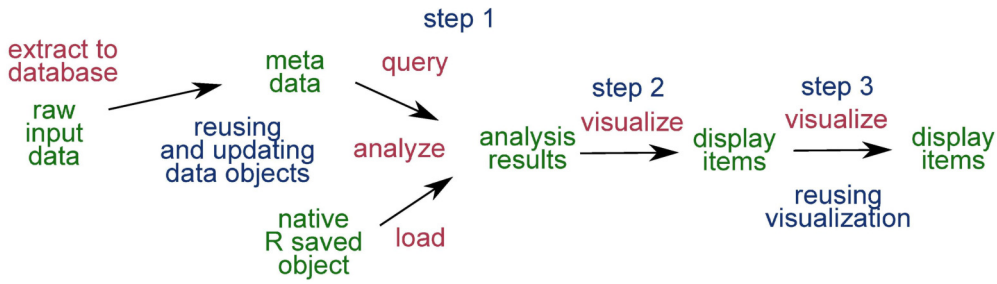


Fig 4. Using Lir to integrate two independent data sets and visualize the results. A: The display items generated by the analysis outlined in (B). The result of each step, represented by the corresponding display item, is taken into consideration when formulating the next question and designing the analysis. **B**: Diagram demonstrating the data flow for an analysis that incorporates a new data set. The new data set associates genes with a measure of the certainty that they are targeted by the same micro-RNA as a gene of interest. The existing relational database object was updated (not re-generated) to include the additional data. The analysis was done in three consecutive steps: First, a reasonable cut-off for the prediction certainty of

the micro-RNA targets was determined (Step 1); then the genes associated to the gene of interest were found (Step 2); then, the mRNA levels of the most interesting of the associated genes was plotted against the mRNA levels of the gene of interest (Step 3). The visualization from Fig 2B was reused in the last step. Data objects are colored in green, data transformations are colored in red, and arrows represent the dependencies.

doi:10.1371/journal.pone.0164023.g004

and *SNX22* (Sorting nexin-22). The Poliovirus receptor protein might provide tumors with a mechanism of immunoevasion, and it plays a role in mediating tumor cell invasion and migration [44]. Sorting nexin-22 may be involved in several stages of intracellular trafficking, inferred from sequence similarity, and contains binding sites for phosphatidylinositol 3-phosphate [45].

The contingency tables of the amplification status and the scatter plots of the mRNA levels of *LAPTM4B* and the two genes are shown in Fig 4A, Panel III. They allow for the following observations: the genes are clearly correlated on the mRNA level, despite the fact that they are not co-amplified. Thus, our data suggest that the mRNA levels of these genes are regulated by miRNA-188, together with *LAPTM4B* mRNA. The phenomenon of one miRNA targeting a complex of functionally related proteins is known [46, 47]. It is therefore possible that *LAPTM4B*, *PVR*, and *SNX22* contribute to shared functions.

Fig 4B outlines the data flow of the combined analyses performed. Existing intermediate results are not regenerated. For example, only the miRNA target data is inserted into the existing relational database, while the already present amplification and mRNA data object is used as it is. This minimizes the time for generating and visualizing new results, thus encouraging an iterative, exploratory approach to data analysis without sacrificing repeatability. In addition, code chunks support ad hoc reuse of code (S1 File, lines 1357–78). In most common use cases achieving code reuse with R's package system is more complex and time-consuming. The use of Lir does not prevent us from using R's package system or the corresponding code reuse paradigm of any other programming platform. Rather, Lir facilitates a systematic approach to organizing computer code that is independent of any particular tool and can be used only if deemed beneficial.

Conclusions

In this study, we have introduced a tool for reproducible computing called Lir. We used Lir to analyze a heterogeneous data set in order to see whether there is putative coregulation between a set of endosomal trafficking regulators. In the analysis, we combined a relational database, an array of data manipulation tools, and a statistical analysis environment. Our results revealed a coamplification of the cancer and lipid transport related genes *NDRG1* and *LAPTM4B*, as well as new genes potentially co-regulated and cooperating with *LAPTM4B*.

The major contribution of Lir within the context of reproducibility is to demonstrate that it is possible and very advantageous to fully document and automate a work flow utilizing a combination of software tools. Using the best tool for each task reduces the total amount of code, thus reducing the opportunity for mistakes, and the amount of invested time [48, 49]. All results and the computer code that generated them are presented as a human-readable document. This document serves two equally important purposes during the development. First, it organizes and presents the intermediate results: the data analysis can be conducted in an iterative, exploratory fashion while faithfully documenting all steps. Second, it organizes, documents and presents all computer executable code: other scientists can inspect the analysis and verify the results. In combination with a version control system, Lir directly facilitates the implementation of best-practice guidelines as delineated in [4–6].

Supporting Information

S1 File. The complete data analysis. This is a plain text file that uses extended markdown as understood by Pandoc. The file is a valid `noweb` source file. It is available at github.com/borisvassilev/endobrca.
(LIR)

S2 File. The generated human-readable document. The final document is an HTML web page that can be viewed with a web browser. It is a faithful representation of the literate source file, which has additionally been cross-linked and prettified. It is available at github.com/borisvassilev/endobrca.
(HTML)

Acknowledgments

We thank Kecheng Zhou for discussions on the topic of miRNA gene regulation. We thank Erkka Valo for providing valuable feedback on the design of the method. The results published here are in part based upon data generated by TCGA managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

Author Contributions

Conceptualization: BV EI.

Data curation: BV RL.

Formal analysis: RL.

Funding acquisition: BV EI SH.

Investigation: BV RL.

Methodology: BV EI.

Resources: EI SH.

Software: BV.

Supervision: EI SH.

Validation: BV RL EI SH.

Visualization: BV.

Writing – original draft: BV RL EI SH.

Writing – review & editing: BV RL EI SH.

References

1. Baggerly K. Disclose all data in publications. Nature. 2010 Sep; 467(7314):401. Available from: http://dx.doi.org/10.1038/467401b. PMID: 20864982
2. Peng RD. Reproducible research in computational science. Science. 2011 Dec; 334(6060):1226–1227. Available from: http://dx.doi.org/10.1126/science.1213847. PMID: 22144613
3. Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, et al. Repeatability of published microarray gene expression analyses. Nat Genet. 2009 Feb; 41(2):149–155. Available from: http://dx.doi.org/10.1038/ng.295. PMID: 19174838

4. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013 Oct; 9(10):e1003285. Available from: http://dx.doi.org/10.1371/journal.pcbi.1003285. PMID: 24204232
5. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. Best practices for scientific computing. PLoS Biol. 2014 Jan; 12(1):e1001745. Available from: http://dx.doi.org/10.1371/journal.pbio.1001745. PMID: 24415924
6. Shade A, Teal TK. Computing Workflows for Biologists: A Roadmap. PLoS Biol. 2015 Nov; 13(11): e1002303. Available from: http://dx.doi.org/10.1371/journal.pbio.1002303. PMID: 26600012
7. Leisch F. Sweave: Dynamic generation of statistical reports using literate data analysis. In: Compstat. Springer; 2002. p. 575–580 doi: 10.1007/978-3-642-57489-4_89
8. Xie Y. knitr: a comprehensive tool for reproducible research in R. Implementing Reproducible Research. 2014;p. 1.
9. Noble WS. A quick guide to organizing computational biology projects. PLoS Comput Biol. 2009 Jul; 5 (7):e1000424. Available from: http://dx.doi.org/10.1371/journal.pcbi.1000424. PMID: 19649301
10. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014. Available from: http://www.R-project.org/.
11. Pérez F, Granger BE. IPython: a System for Interactive Scientific Computing. Computing in Science and Engineering. 2007 May; 9(3):21–29. Available from: http://ipython.org.
12. Shen H. Interactive notebooks: Sharing the code. Nature. 2014 Nov; 515(7525):151–152. Available from: http://dx.doi.org/10.1038/515151a. PMID: 25373681
13. Ovaska K, Laakso M, Haapa-Paananen S, Louhimo R, Chen P, Aittomäki V, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. Genome Med. 2010; 2(9):65. doi: 10.1186/gm186 PMID: 20822536
14. Börnigen D, Moon YS, Rahnavard G, Waldron L, McIver L, Shafquat A, et al. A reproducible approach to high-throughput biological data acquisition and integration. PeerJ. 2015; 3:e791. Available from: http://dx.doi.org/10.7717/peerj.791. PMID: 26157642
15. Kitchin JR. Examples of Effective Data Sharing in Scientific Publishing. ACS Catalysis. 2015; 5 (6):3894–3899. Available from: http://dx.doi.org/10.1021/acscatal.5b00538.
16. Vassilev B. Lir; 2015. Available from: https://github.com/borisvassilev/lir.
17. Knuth DE. Literate programming. The Computer Journal. 1984; 27(2):97–111. Available from: http://comjnl.oxfordjournals.org/content/27/2/97.short.
18. Network TCGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct; 490 (7418):61–70. Available from: http://dx.doi.org/10.1038/nature11412. PMID: 23000897
19. Ramsey N. Literate programming simplified. IEEE software. 1994; 11(5):97–105. Available from: http://www.computer.org/csdl/mags/so/1994/05/s5097.pdf. doi: 10.1109/52.311070
20. Wielemaker J, Schrijvers T, Triska M, Lager T. SWI-Prolog. Theory and Practice of Logic Programming. 2012; 12(1–2):67–96. doi: 10.1017/S1471068411000494
21. Carvalho B, Bengtsson H, Speed TP, Irizarry RA. Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data. Biostatistics. 2007 Apr; 8(2):485–499. Available from: http://dx.doi.org/10.1093/biostatistics/kxl042. PMID: 17189563
22. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004 Oct; 5(4):557–572. Available from: http://dx.doi.org/10.1093/biostatistics/kxh008. PMID: 15475419
23. Therneau Terry M, Grambsch Patricia M. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.
24. Therneau TM. A Package for Survival Analysis in S; 2014. R package version 2.37-7. Available from: http://CRAN.R-project.org/package=survival.
25. Vassilev B, Sihto H, Li S, Hölttä-Vuori M, Ilola J, Lundin J, et al. Elevated levels of StAR-related lipid transfer protein 3 alter cholesterol balance and adhesiveness of breast cancer cells: potential mechanisms contributing to progression of HER2-positive breast cancers. Am J Pathol. 2015 Apr; 185 (4):987–1000. Available from: http://dx.doi.org/10.1016/j.ajpath.2014.12.018. PMID: 25681734
26. Fang BA, Kovačević Ž, Park KC, Kalinowski DS, Jansson PJ, Lane DJR, et al. Molecular functions of the iron-regulated metastasis suppressor, NDRG1, and its potential as a molecular target for cancer therapy. Biochim Biophys Acta. 2014 Jan; 1845(1):1–19. Available from: http://dx.doi.org/10.1016/j.bbcan.2013.11.002. PMID: 24269900
27. Pietiäinen V, Vassilev B, Blom T, Wang W, Nelson J, Bittman R, et al. NDRG1 functions in LDL receptor trafficking by regulating endosomal recycling and degradation. J Cell Sci. 2013 Sep; 126(Pt 17):3961–3971. Available from: http://dx.doi.org/10.1242/jcs.128132. PMID: 23813961

28. Blom T, Li S, Dichlberger A, Bäck N, Kim YA, Loizides-Mangold U, et al. LAPTM4B facilitates late endosomal ceramide export to control cell death pathways. Nat Chem Biol. 2015 Oct; 11(10):799–806. Available from: http://dx.doi.org/10.1038/nchembio.1889. PMID: 26280656

29. Kauraniemi P, Kallioniemi A. Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. Endocr Relat Cancer. 2006 Mar; 13(1):39–49. Available from: http://dx.doi.org/10.1677/erc.1.01147. PMID: 16601278

30. Parris TZ, Kovács A, Hajizadeh S, Nemes S, Semaan M, Levin M, et al. Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. Oncogenesis. 2014; 3:e95. Available from: http://dx.doi.org/10.1038/oncsis.2014.8. PMID: 24662924

31. Li Y, Zou L, Li Q, Haibe-Kains B, Tian R, Li Y, et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. Nat Med. 2010 Feb; 16(2):214–218. Available from: http://dx.doi.org/10.1038/nm.2090. PMID: 20098429

32. Staaf J, Jönsson G, Ringnér M, Vallon-Christersson J, Grabau D, Arason A, et al. High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. Breast Cancer Res. 2010; 12(3):R25. Available from: http://dx.doi.org/10.1186/bcr2568. PMID: 20459607

33. Chan HJ, Li H, Liu Z, Yuan YC, Mortimer J, Chen S. SERPINA1 is a direct estrogen receptor target gene and a predictor of survival in breast cancer patients. Oncotarget. 2015 Sep; 6(28):25815–25827. Available from: http://dx.doi.org/10.18632/oncotarget.4441. PMID: 26158350

34. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan; 116(2):281–297. doi: 10.1016/S0092-8674(04)00045-5 PMID: 14744438

35. Ambros V. The functions of animal microRNAs. Nature. 2004 Sep; 431(7006):350–355. Available from: http://dx.doi.org/10.1038/nature02871. PMID: 15372042

36. He XH, Zhu W, Yuan P, Jiang S, Li D, Zhang HW, et al. miR-155 downregulates ErbB2 and suppresses ErbB2-induced malignant transformation of breast epithelial cells. Oncogene. 2016 Apr;Available from: http://dx.doi.org/10.1038/onc.2016.132 PMID: 27065318.

37. Zhang H, Qi S, Zhang T, Wang A, Liu R, Guo J, et al. miR-188-5p inhibits tumour growth and metastasis in prostate cancer by repressing LAPTM4B expression. Oncotarget. 2015 Mar; 6(8):6092–6104. Available from: http://dx.doi.org/10.18632/oncotarget.3341. PMID: 25714029

38. Luo EC, Chang YC, Sher YP, Huang WY, Chuang LL, Chiu YC, et al. MicroRNA-769-3p down-regulates NDRG1 and enhances apoptosis in MCF-7 cells during reoxygenation. Sci Rep. 2014; 4:5908. Available from: http://dx.doi.org/10.1038/srep05908. PMID: 25081069

39. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat Methods. 2015 Aug; 12(8):697. Available from: http://dx.doi.org/10.1038/nmeth.3485. PMID: 26226356

40. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008 Jan; 36(Database issue):D149–D153. Available from: http://dx.doi.org/10.1093/nar/gkm995. PMID: 18158296

41. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015 Jan; 43(Database issue):D146–D152. Available from: http://dx.doi.org/10.1093/nar/gku1104. PMID: 25378301

42. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell. 2006 Sep; 126(6):1203–1217. Available from: http://dx.doi.org/10.1016/j.cell.2006.07.031. PMID: 16990141

43. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015; 4. Available from: http://dx.doi.org/10.7554/eLife.05005 PMID: 26267216.

44. Sloan KE, Eustace BK, Stewart JK, Zehetmeier C, Torella C, Simeone M, et al. CD155/PVR plays a key role in cell motility during tumor cell invasion and migration. BMC Cancer. 2004 Oct; 4:73. Available from: http://dx.doi.org/10.1186/1471-2407-4-73. PMID: 15471548

45. Song J, Zhao KQ, Newman CLL, Vinarov DA, Markley JL. Solution structure of human sorting nexin 22. Protein Sci. 2007 May; 16(5):807–814. Available from: http://dx.doi.org/10.1110/ps.072752407. PMID: 17400918

46. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol. 2004 Nov; 2(11):e363. Available from: http://dx.doi.org/10.1371/journal.pbio.0020363. PMID: 15502875

47. Didiano D, Hobert O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nat Struct Mol Biol. 2006 Sep; 13(9):849–851. Available from: http://dx.doi.org/10.1038/nsmb1138. PMID: 16921378

48. Prechelt L. An empirical comparison of seven programming languages. Computer. 2000; 33(10):23–29. doi: 10.1109/2.876288

49. McConnell S. Code complete. Pearson Education; 2004.